

=> d que

L4 72 SEA FILE=HCAPLUS ABB=ON PLU=ON ("BARRY C"/AU OR "BARRY C C"/AU OR "BARRY C D"/AU OR "BARRY C DAVID"/AU OR "BARRY C E"/AU OR "BARRY C E III"/AU OR "BARRY C J"/AU OR "BARRY C L"/AU OR "BARRY C M"/AU OR "BARRY C M F"/AU OR "BARRY C MARVIN"/AU OR "BARRY C N"/AU OR "BARRY C P"/AU OR "BARRY C S"/AU OR "BARRY CAROLINE"/AU OR "BARRY CAROLYN B"/AU)

L5 126 SEA FILE=HCAPLUS ABB=ON PLU=ON ("BOUGUELERET L"/AU OR "BOUGUELERET LYDIE"/AU)

L6 130 SEA FILE=HCAPLUS ABB=ON PLU=ON ("CHUMAKOV I"/AU OR "CHUMAKOV I F"/AU OR "CHUMAKOV I M"/AU OR "CHUMAKOV I R"/AU OR "CHUMAKOV I S"/AU OR "CHUMAKOV I V"/AU OR "CHUMAKOV ILYA"/AU OR "CHUMAKOV ILYA M"/AU)

L7 12 SEA FILE=HCAPLUS ABB=ON PLU=ON (L*** OR L*** OR L***)

L8 27 SEA FILE=HCAPLUS ABB=ON PLU=ON (L4 AND (L5 OR L6 OR L7)) OR (L5 AND (L6 OR L7)) OR (L6 AND L7)

L9 302 SEA FILE=HCAPLUS ABB=ON PLU=ON (L4 OR L5 OR L6 OR L7)

L10 1 SEA FILE=HCAPLUS ABB=ON PLU=ON L9 AND BAP?

L11 23 SEA FILE=HCAPLUS ABB=ON PLU=ON L9 AND PROSTAT?

L12 37 SEA FILE=HCAPLUS ABB=ON PLU=ON L8 OR (L10 OR L11)

=> d ibib abs 1-37

L12 ANSWER 1 OF 37 HCAPLUS COPYRIGHT 2006 ACS on STN

ACCESSION NUMBER: 2006:367166 HCAPLUS

DOCUMENT NUMBER: 144:405904

TITLE: Polynucleotides encoding human TBC-1 protein, polymorphic markers therein, and uses thereof

INVENTOR(S): Blumenfeld, Marta; **Bougueleret, Lydie;**
Chumakov, Ilya

PATENT ASSIGNEE(S): Genset S.A., Fr.

SOURCE: U.S. Pat. Appl. Publ., 114 pp.

CODEN: USXXCO

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2006084073	A1	20060420	US 2004-970750	20041020
PRIORITY APPLN. INFO.:			US 2004-970750	20041020

AB The invention concerns genomic and cDNA sequences of the human TBC-1 gene. The invention also concerns polypeptides encoded by the TBC-1 gene. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the TBC-1 gene useful in genetic anal. The TBC-1 gene is located in a candidate region for **prostate** cancer on chromosome 4. Expression of TBC-1 gene leads to production of at least two mRNA mols. which have different first exons, exon 1 and 1bis, due to alternative splicing.

L12 ANSWER 2 OF 37 HCAPLUS COPYRIGHT 2006 ACS on STN

ACCESSION NUMBER: 2005:182920 HCAPLUS

DOCUMENT NUMBER: 142:258503

TITLE: Secreted polypeptide species in human plasma, detection assays for smaller proteins and tryptic peptides, and expression profiles useful for disease

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:42:00 ; Search time 39.2894 Seconds
(without alignments)
3989.295 Million cell updates/sec

Title: us-09-603-665-5_COPY_1_1629
Perfect score: 8212
Sequence: 1 MTSLAQQLRLALPQSDASL.....NPLPSVRRKALDILNNKLLQ 1629

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	673.5	8.2	1649	T39938	hypothetical prote
2	654	8.0	1769	S53378	probable membrane
3	536	6.5	1650	T27864	hypothetical prote
4	248	3.0	2670	T37919	GCN1 homolog - fis
5	241	2.9	2895	H85362	hypothetical prote
6	229	2.8	2059	T41933	large tegument pro
7	228	2.8	3433	S28381	urothrin - human
8	228	2.8	4131	T21085	hypothetical prote
9	227.5	2.8	2954	T14156	kinesin-related pr
10	226	2.8	1447	F82909	hypothetical prote
11	226	2.8	1830	E82969	conserved hypochet
12	224	2.7	2297	AB2494	hypothetical prote
13	219.5	2.7	2819	A90551	conserved hypochet
14	215	2.6	2663	S28261	centromere protein
15	212.5	2.6	4385	T29042	hypothetical prote
16	212	2.6	2108	S28417	CDC39 protein - ye
17	210.5	2.6	2269	T28677	thoptry protein -
18	209	2.5	2401	T28676	thoptry protein -
19	208	2.5	4717	T41581	hypothetical colle
20	207.5	2.5	1655	S47446	nucleoporin Np188
21	207.5	2.5	4859	S74173	ryanodine receptor
22	207	2.5	3744	S46715	hypothetical prote
23	206	2.5	1919	T40032	hypothetical prote
24	204	2.5	3187	JC5837	364k Golgi complex
25	202.5	2.5	2493	S45734	probable membrane
26	202	2.5	3079	RGBY12	probable GTPase-ac
27	201.5	2.5	2672	A48126	translational activa
28	201	2.4	1957	T38077	hypothetical colle
29	199.5	2.4	1964	A59282	nonmuscle myosin I

30	198.5	2.4	1937	2	T41023	probable nuclear p
31	198.5	2.4	3795	2	T00831	hypothetical prote
32	198	2.4	2453	2	T40540	hypothetical prote
33	197.5	2.4	1922	2	T00637	hypothetical prote
34	197.5	2.4	2094	2	S33124	cpf protein - huma
35	197.5	2.4	2166	2	G70163	hypothetical prote
36	196	2.4	1690	2	T13030	microtubule bindin
37	195.5	2.4	1420	2	S54471	probable membrane
38	194.5	2.4	2376	2	S48405	probable membrane
39	194.5	2.4	4872	2	S27272	ryanodine receptor
40	194.5	2.4	4924	2	T50176	probable peptide s
41	194	2.4	2110	2	H96803	unknown protein T5
42	192	2.3	2470	2	S57085	1-phosphatidylinos
43	192	2.3	2712	2	T05113	hypothetical prote
44	191.5	2.3	1732	2	G84664	hypothetical prote
45	191.5	2.3	3655	2	T38084	TRAP-like protein

ALIGNMENTS

RESULT 1									
T39938									
hypothetical protein SPC23B6.04c - fission yeast (Schizosaccharomyces pombe)									
C:Species: Schizosaccharomyces pombe									
C>Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text_change 09-Jul-2004									
C:Accession: T39938									
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Pohl, T.									
Submitted to the EMBL Data Library, May 1998									
A:Reference number: Z21892									
A:Accession: T39938									
A:Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: DNA									
A:Residues: 1-1649 <MOO>									
A:Cross-references: UNIPROT:O60179; UNIPARC:UP1000013AF93; EMBL:AL023287; PDB:CAA18872.1									
A:Experimental source: strain 972h-; cosmid c23B6									
C:Genetics:									
A:Gene: SPDB:SPBC23B6.04c									
A:Map position: 2									
Query Match									
Best Local Similarity 20.5%; Pred. No. 1.7e-25;									
Matches 348; Conservative 263; Mismatches 490; Indels 593; Gaps 61;									
QY	2	TSLSAQQLRLALPQSDASLSRDEV--ASLLPDKREAAITIDBDTAFALIGCTGLEELGI	58						
DB	3	SSLSQQLKNI---QSN-NVLKINKIRAPSLLDPKVAADMDEEITVTAVSGFELAVH	58						
QY	59	DPSPEQFAPLFSQAKTLERSVQTKAVNKQDENISLFLHSPYFLKPAQCKLEMLI	118						
DB	59	EPRLITFEKTLLEBOSVQVRVLNRTENKIDECVQILRLAPFTEGNALKVLEMLI	118						
QY	119	HRFHILYNODSLIACVLPYHETRIFFRVITQLKINKSKRMFWLDPVKGSPVLAGTIL	178						
DB	119	RRRSIHETVSDFFLSFLPHDHPFARILGCSK-PKSRPDLFLLENAIKMP-VLISRADI	176						
QY	179	ITHCYKDLGMDTICSLVTSKVFAPFPGSSAQLRVLLAFYSTIYSAVAADVDN-	237						
DB	177	VHLSLRDKPEFAPFAQFQNTASHNYP-----ELAFPMGTMEVLAHMSNEDP	229						
QY	238	--IIAKLPYIQGLK--SLSPYRAATYMIQISVKVMENFVSLASQILKTLTKI	293						
DB	230	NVLDRFLKAVSAVSVSSI-DQIAGFMILSSIASLSPSIIIPVLSAIDRLSP-	287						
QY	294	PSLTKDGLSLIVLL-----QKPESTLG-----	317						
DB	288	DNMKPALICVGHLPQSCSFEDDEQLKESGASSLIELSQEHLDEFFVSYVWSL	346						
QY	318	-----KKPFHLNCPVPLITILHGISETYDVSPPLRLPVLVYSIIHHVTGETEGMDG	372						
DB	347	IKSRQKQKRRLLSLD--TSISQIRVTHQAKFLSLVIP-----VNQDFKAL--	392						
QY	373	QIVRHLEALITKISLKNLIDHLASLLFEEYISYSQSEMDSNKSVSLNEQFLPLIRLL	432						

D	393	QSYRILDSV	QPERKEGKDLNLT	-----	ODKKSSTFSKKDEVL	-----	436
Q	433	ESKYPRILDV	LEELHKEIADLKQOE	LFHOVS	STSGKYQ	FLADSDTSLM	SLHPLA 492
D	437	-----	-----	-----	-----	-----	-----
Q	493	PVRLANMHL	KKIKNTS	KEGVDS	FEICAV	LARLDODDIDV	LAI
D	471	-----	-----	-----	-----	-----	-----
Q	553	TISNLT	FORAELSKNGEWEY	KIADIL	IKELI	SENDOLSNQV	CLPEVV -IN 610
D	493	IILSLIEL	-----	-----	-----	-----	-----
Q	611	NDPES	AMKAIYLSKSG	ICSLHPLRG	EBALBNYIK	STPKGLIG	ANQKMIELAD 670
D	525	SKDTEVSR	ALNLT	ILT	-----	-----	-----
Q	671	NINQD	PSMUKW	VEDLIS	VEESFN	KQVTFHVL	SLVSCSS
D	548	NL	-----	-----	-----	-----	-----
Q	731	SLQKIK	IKLES	VTAVEI	PSEWHI	ELMDRG	IPVELMAHYEELN
D	563	-----	-----	-----	-----	-----	-----
Q	791	FSLK	KFIYALKAP	SPFKGDI	WMNPEOL	KEBSRYIHL	IGLFEM
D	597	SYLPER	FTEKPK	KNASK	-----	-----	-----
Q	851	LF	IKHLEDV	FOLFK	CSVLMT	WYSSLS	NPLNC
D	620	SPLSH	-----	-----	-----	-----	-----
Q	911	QLAS	ISSPV	TSLIN	GS	PVKEVR	RAAIQCLQ
D	634	-----	-----	-----	-----	-----	-----
Q	971	AYV	ODATL	FEELQ	REKELK	SHOKL	SETLNL
D	652	-----	-----	-----	-----	-----	-----
Q	1031	LSQ	LPAEOL	LEKIQE	PTAVAL	KDEAMVHL	TLGK
D	705	LSQAI	-----	-----	-----	-----	-----
Q	1091	TTKEL	VAGMP	IOITALE	KTKPF	FPAIIS	DEKVOQLRML
D	734	SODE	-----	-----	-----	-----	-----
Q	1151	FKGIS	VNAEO	VRLE	BPDKAK	PLGTVO	QKRQKMO
D	756	FK	-----	-----	-----	-----	-----
Q	1211	LELLO	HKKKUR	POILVPT	FNLSR	CLEP	POEGMNEYKOLIS
D	795	LELE	ETKAAS	YPKLAS	P	LEVANSY	ALKE
Q	1266	LS	PDGKI	PKDIL	DEKFN	VELI	QICIR
D	853	LSF	-----	-----	-----	-----	-----
Q	1326	IF	TMGAN	WR	LDDTYS	FOVINK	VKVIPAL
D	900	IF	TMGS	TSR	DAFS	IHVLEQ	TVKVIS
Q	1384	VDAL	PH	REHRL	ILVOL	VDTL	TA
D	948	VNA	FP	PHI	PO	RRLALY	HLVLOT
Q	1444	TE	FW	SVC	CE	BS	HO

```

Db      999 HFCCLITVQSFVAVRIGISINOCNRFCLKSLBEOGNSDNGKAVSLIKLDELPE-----MD 1053
QY      1502 VETHTSKOLRHKFKFLSVSMQOLLSSNNF---LKKVSGGPEILKGLERLLLETVLGYI 1558
Db      1054 VDLATLGLSLR-KVLEL--ISLVSRAKNFAPDLAKIMENSVSVFE-IQAGLPESIKLLI 1109
QY      1559 SAVAGS---MEENNAKLTATKFPARALLSKAYDDLDKNNALLPTEFTPIVIRGLGNPLPSV 1615
Db      1110 TLSQGSNEME-----LGHVYVALRSYIHLPLNELFCTVLGKLHLDERALL 1155
QY      1616 RKKALDLNNKLOQ 1629
Db      1156 RKKALSTVQORVQO 1169

RESULT 2
SS3378
probable membrane protein YJH109c - yeast (Saccharomyces cerevisiae)
N.Alternate names: hypothetical protein J0808
C.Species: Saccharomyces cerevisiae
C.Date: 05-May-1995 #sequence revision 01-Sep-1995 #text_change 09-Jul-2004
C.Accession: S53378; S56887; S57359
R.Rasmussen, S.W.
submitted to the EMBL Data Library, February 1995
A.Description: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and
A.Reference number: S53376
A.Accession: S53378
A:Molecule type: DNA
A:Residues: 1-1769 <RAM>
A:Cross-references: UNIPROT:P42945; UNIPARC:UPI0000052F40; EMBL:X65021; NID:g728698; PID
R.Rasmussen, S.W.
submitted to the Protein Sequence Database, September 1995
A.Reference number: S56876
A.Accession: S56887
A:Molecule type: DNA
A:Residues: 1-1769 <RAM>
A:Cross-references: UNIPARC:UPI0000052F40; EMBL:Z49384; NID:g1008292; PID:g1008293; MIPS
R.Rasmussen, S.W.
Yeast 11, 873-883, 1995
A.Title: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and CSD3 ge
A.Reference number: S57357; MUID:96090136; PMID:7483851
A.Accession: S57359
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1769 <RAM>
A:Cross-references: UNIPARC:UPI0000052F40; EMBL:X65021; NID:g728698; PID:CA59385.1; PIR
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
C.Genetics:
A:Cross-references: SGD:S0003645
A:Map position: 10L
C:Keywords: transmembrane protein

Query Match      8.0%; Score 654; DB 2; Length 1769;
Best Local Similarity 20.8%; Pred.No.1.7e-24;
Matches 366; Conservative 252; Mismatches 542; Indels 598; Gaps 65;

1 MTSLAQOLRLALPQSDASL-----LSRDEVASLLPPEKCAATIDRDTAFAGTGLEELL 56
Db      1 MSSLSDLQAAVSNNAVTALDKRROKLSASLIYNSKTAATDYPFIPENASKALEELS 60
QY      57 GIDPFEQFEAPLFSQALKTERSVQTKAVNKOLDENISLFLHLSPPYLLKPAQKCLEW 116
Db      61 QIEPFAIFSRFLPSESSISLDRNVQTKKEIKDLDAINAVYLLASQWYLAFTLHATEW 120
QY      117 LIHRPHILHYNODSLIACVLPYHETRIFFVRVQLKINNSKRWFWLLP-----VK 167
Db      121 LVRRQIHVKNTKEMLLSTLNTVYQTPVFKIILSTIK-----LPPLFNCLSNFR 169
QY      168 QSGVPLAGTLITHCYKDLGFMDPFCISLVTKSVKVPFAEYFGSSAQLRVLLAFYASTIVSA 227
Db      170 SEKPFTAL-TWI-KLFNMDMDFLKYTSYLDQCCKHNATYTNQ-----LLFTCCFIN- 219
QY      228 LVAAEDVSDNITAKLPPI-----QKGKSSLDPYRAATYMIQGISVKTMENTFVNSLA 283

```

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:41:34 ; Search time 166.656 Seconds
(without alignments)
9041.668 Million cell updates/sec

Title: US-09-603-665-5_COPY_1_1629
Perfect score: 8212
Sequence: 1 MTSIAQQLRLALPQSDASL.....NPLPSYRRKALDLINKLQO 1629

Scoring table: BLOSUM62
Gapop 10.0 ; Gapext 0.5

Searched: 2849598 segs, 925015592 residues
Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 7.2:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8204	99.9	2144	1 HEATR_HUMAN	09h583 homo sapien
2	7755.5	94.4	2063	05t307_HUMAN	05t307 homo sapien
3	4891	59.6	1180	03JUNW7_MOUSE	03jun7 mus musculu
4	4516	55.0	1090	03TNEI1_MOUSE	03tnei1 mus musculu
5	4118	50.1	2159	07SY48_BRARE	07sy48 brachydanio
6	4072	49.6	984	03TWK0_MOUSE	03twk0 mus musculu
7	3589	43.7	2288	04T723_TETNG	04t723 tetradon n
8	3099	37.7	733	08BLJ4_MOUSE	08blj4 mus musculu
9	3095	37.7	733	03T910_MOUSE	03t910 mus musculu
10	2973	36.2	1106	06P197_HUMAN	06p197 homo sapien
11	2298	28.0	1378	07T153_BRARE	07t153 brachydanio
12	2298	28.0	1336	07T152_BRARE	07t152 brachydanio
13	2181.5	26.6	958	1 HEATR_MACFA	09gm44 macaca fasc
14	1916	23.3	897	1 HEATR_HUMAN	08n717 homo sapien
15	1444	17.6	2104	07PND6_ANOGA	07pnd6 anopheles g
16	1398.5	17.0	2096	06AWS0_DROME	06aws0 drosophila
17	1396.5	17.0	2096	1 HEATR_DROME	09vm75 drosophila
18	948.5	11.6	743	03VIX6_MOUSE	03vix6 mus musculu
19	852	10.4	168	2 Q2KIF9_BOVIN	02kif9 bos taurus
20	832	10.1	2251	2 Q4P937_USUMA	04p937 ussilaogo ma
21	809.5	9.9	1690	2 Q8T9E7_DROME	08t9e7 drosophila
22	799.5	9.7	2237	2 Q54WU4_DICDI	054w44 dicystoteli
23	786.5	9.6	1830	1 HEATR_ARATH	09c824 arabidopsis
24	774	9.4	2021	2 Q5K8Q2_CRYNE	05k8q2 cryptococcu
25	767	9.3	2021	2 Q55NB7_CRYNE	055nb7 cryptococcu
26	763.5	9.3	1857	2 Q6BXO6_DEBHA	06bx06 debryomyc
27	758.5	9.2	1801	2 Q5CAF8_EMENI	05caf8 aspergillus
28	757.5	9.2	2122	2 Q5CAF8_ORISA	05caf8 oryza sativ
29	715.5	8.7	1802	2 Q2ULC6_ASPOR	02ulc6 aspergillus
30	707.5	8.6	1770	2 Q6F793_CANCA	06f793 candida gla
31	707.5	8.6	1798	2 Q4WLI9_ASPFU	04wli9 aspergillus

32	700	8.5	1788	2 Q7R2M6_NEUCR	07r2m6 neurospora
33	696.5	8.5	1774	2 Q6CJ57_KLUIA	06cj57 kluyveromyc
34	684.5	8.3	1818	2 Q592X6_CANAL	0592x6 candida alb
35	682	8.3	1774	2 Q754J8_ASHAO	0754j8 ashyia goss
36	673.5	8.2	1649	1 YG44_SCHPO	060179 schizosacch
37	654	8.0	1769	1 UTP10_YEAST	042945 saccharomyc
38	641	7.8	1635	2 Q6C457_YARLI	06c457 yarrowia li
39	629.5	7.7	1806	2 Q41696_GIBZE	041696 gibberella
40	600	7.3	120	2 Q6P664_HUMAN	06p664 homo sapien
41	556.5	6.8	1660	2 Q61Y68_CAEER	061y68 caenorhabdi
42	536	6.5	1650	1 HEATR_CAEER	023495 caenorhabdi
43	405.5	4.9	2034	2 Q519S9_ENTHI	0519s9 entamoeba h
44	334	4.1	2606	2 Q4DAC4_TRYCR	04dac4 trypanosoma
45	330.5	4.0	2631	2 Q3BPN3_YITRP	03bpn3 trypanosoma

ALIGNMENTS

RESULT 1
HEATR_HUMAN
ID HEATR_HUMAN STANDARD; PRT; 2144 AA.
AC 09H583; 05T307; 09NM23;
PT 01-JUN-2001; integrated into UniProtKB/Swiss-Prot.
DT 19-JUL-2005; sequence version 3.
DT 07-MAR-2006; entry version 33.
DE HEAT repeat-containing protein 1 (Protein BAP28).
GN Name=HEATR1; Synonyms=BAP28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE AND VARIANT-9-SER-1694; ALA-1854; ASP-1967 AND
RP GLY-2017.
RA Bouguetel L., Chumakov V., Barry O., Cohen-Akenine A.;
RT "A novel BAP28 gene and protein".
RL Patent number WO0106697, 04-JAN-2001.
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RG Human chromosome 1 international sequencing consortium;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RX PubMed=14702039; DOI=10.1038/ng1285.
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 1777-2144.
RX Published=14702039; DOI=10.1038/ng1285.
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi N., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimiya K., Ishihashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tani H., Kimeta M., Watanabe M., Hirao K., Ichida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fuji A., Hara H., Tanase T.-O., Nomura Y.,
RA Togiwa S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuma S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Saito S.,
RA Moriya S., Montiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamaguchi M., Watanabe K., Kumagai A., Itakura S., Fukuzawa Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura T., Nakajima Y., Mizuno T., Morinaga M., Komatsu T.,
RA Togoashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:46:09 ; Search time 28.4956 Seconds
(without alignments)
5003.834 Million cell updates/sec

Title: US-09-603-665-5_COPY_1_1629
Perfect score: 8212
Sequence: 1 MTSLAQQLQRLALPQSDASL.....NPUPSVRRKALDLNNKLAQQ 1629

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues
Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /EMC_Celerra_SIDS3/prodata/2/1aa/5_COMB.pep:*
- 2: /EMC_Celerra_SIDS3/prodata/2/1aa/6_COMB.pep:*
- 3: /EMC_Celerra_SIDS3/prodata/2/1aa/7_COMB.pep:*
- 4: /EMC_Celerra_SIDS3/prodata/2/1aa/H_COMB.pep:*
- 5: /EMC_Celerra_SIDS3/prodata/2/1aa/PCTUS_COMB.pep:*
- 6: /EMC_Celerra_SIDS3/prodata/2/1aa/RE_COMB.pep:*
- 7: /EMC_Celerra_SIDS3/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	310	3.8	200	2 US-09-248-796A-15108	Sequence 15108, A
2	228	2.8	3433	2 US-09-091-501B-10	Sequence 10, Appl
3	228	2.8	3433	2 US-09-538-092-1136	Sequence 1136, Ap
4	227.5	2.8	2954	2 US-09-150-867-1	Sequence 1, Appl1
5	227.5	2.8	2954	3 US-09-724-584-1	Sequence 1, Appl1
6	222.5	2.7	3830	2 US-09-693-205A-4	Sequence 4, Appl1
7	216.5	2.6	3829	2 US-09-693-205A-16	Sequence 16, Appl
8	215	2.6	2663	2 US-09-538-092-1252	Sequence 1252, Ap
9	214.5	2.6	3829	2 US-09-693-205A-2	Sequence 2, Appl1
10	212.5	2.6	2662	2 US-09-595-684B-31	Sequence 31, Appl
11	212	2.6	2108	2 US-09-538-092-87	Sequence 87, Appl
12	206.5	2.5	186	2 US-08-973-462-8	Sequence 8, Appl1
13	206.5	2.5	2482	1 US-08-328-254-6	Sequence 6, Appl1
14	206.5	2.5	3448	1 US-08-353-700-1	Sequence 1, Appl1
15	206.5	2.5	3248	5 PCT-US95-16216-1	Sequence 1, Appl1
16	203.5	2.5	3210	2 US-09-538-092-1154	Sequence 1154, Ap
17	202	2.5	3066	2 US-08-952-127-12	Sequence 12, Appl
18	201.5	2.5	1979	2 US-09-949-016-6468	Sequence 6468, Ap
19	201.5	2.5	2047	2 US-09-949-016-7404	Sequence 7404, Ap
20	201.5	2.5	10182	2 US-09-134-001C-3159	Sequence 3159, Ap
21	198	2.4	3079	5 PCT-US94-00198-4	Sequence 4, Appl1
22	197.5	2.4	2186	2 US-09-949-016-10828	Sequence 10828, A
23	197.5	2.4	2349	2 US-09-538-092-914	Sequence 914, Appl
24	194.5	2.4	2316	2 US-09-424-783-3	Sequence 3, Appl1
25	194	2.4	2316	2 US-09-949-016-10280	Sequence 10280, A
26	194	2.4	3056	1 US-08-508-836A-8	Sequence 8, Appl1

27	194	2.4	3056	1 US-08-629-001A-3	Sequence 3, Appl1
28	194	2.4	3056	1 US-08-874-266-2	Sequence 2, Appl1
29	194	2.4	3056	2 US-08-642-274D-3	Sequence 3, Appl1
30	194	2.4	3056	2 US-08-952-127-3	Sequence 3, Appl1
31	194	2.4	3056	2 US-08-952-014C-3	Sequence 3, Appl1
32	194	2.4	3056	2 US-09-360-416-2	Sequence 2, Appl1
33	194	2.4	3056	2 US-08-984-090-2	Sequence 2, Appl1
34	194	2.4	3056	2 US-10-351-733-1	Sequence 1, Appl1
35	194	2.4	3056	2 US-10-175-225-2	Sequence 2, Appl1
36	193	2.4	2470	2 US-08-265-967C-2	Sequence 2, Appl1
37	193	2.4	2470	2 US-08-305-790B-3	Sequence 3, Appl1
38	192.5	2.3	3057	2 US-09-360-416-3	Sequence 3, Appl1
39	192.5	2.3	3057	2 US-10-175-225-3	Sequence 3, Appl1
40	192	2.3	2474	2 US-08-265-967C-3	Sequence 3, Appl1
41	192	2.3	2474	2 US-08-305-790B-4	Sequence 4, Appl1
42	191	2.3	5024	2 US-09-710-279-2964	Sequence 2964, Ap
43	190.5	2.3	1698	2 US-09-315-793-12	Sequence 12, Appl
44	190	2.3	1093	2 US-09-315-793-52	Sequence 52, Appl
45	190	2.3	1093	2 US-09-538-092-701	Sequence 701, Appl

ALIGNMENTS

```

RESULT 1
US-09-248-796A-15108
; Sequence 15108, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15108
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-15108

Query Match      3.8%; Score 310; DB 2; Length 200;
Best Local Similarity 38.0%; Pred. No. 7.8e-18;
Matches 76; Conservative 36; Mismatches 70; Indels 18; Gaps 5;

QY      1 MTSLAQQLQRL-----ALPQSDASL-SRDEVASLFPKRNATIDRDTAFAGCTGL 52
        |||||
DB      4 MSLSAQSLQSLNKTASVALDRKORSKLSR-----SLRDPQQAATQDYEIYEINTEGL 59

QY      53 EELLGIDPFEQFEAPLFSQAKTERSVQYAKANKQKDENISLPIHSPFLKPAQK 112
        |||||
DB      60 EDLCEDSDRFNFKLTFSETSVNLDNRNLOTXDVLSQDKNIDAFLLTGPPYGLTSSLK 119

QY      113 CLEWLIHREHIALYVODSLIACVLPYHETRIFVRYIOLKINNSKRWFWLIPVK-OSGV 171
        |||||
DB      120 AEWLVRRRHAIHNAELMILITALYFQHPVFEVLANIVPKNLPIQFEMLVGYKQDLKT 179

QY      172 PLAKGTLTHCYKDGFMDF 191
        |||||
DB      180 PPASSIL-----KGSAMNF 194

RESULT 2
US-09-091-501B-10
; Sequence 10, Application US/09091501B
; Patent No. 6518413
; GENERAL INFORMATION:
; APPLICANT: Tinsley, Jonathon M

```


APPLICANT: Davies, Kay E
TITLE OF INVENTION: Utrrophin gene expression
FILE REFERENCE: 620-42
CURRENT APPLICATION NUMBER: US/09/091,501B
CURRENT FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: PCT/GB96/03156
PRIOR FILING DATE: 1996-12-19
PRIOR APPLICATION NUMBER: GB 9525962.8
PRIOR FILING DATE: 1995-12-19
PRIOR APPLICATION NUMBER: GB 9615797.9
PRIOR FILING DATE: 1996-07-26
PRIOR APPLICATION NUMBER: GB 9622174.2
PRIOR FILING DATE: 1996-10-24
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 10
LENGTH: 3433
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc feature
LOCATION: (239) ... (250)
OTHER INFORMATION: Description of Artificial Sequence: Full length
OTHER INFORMATION: utrophin construct; Xaa = unknown
US-09-091-501B-10

Query Match 2.8%; Score 228; DB 2; Length 3433;
Best Local Similarity 18.7%; Pred. No. 8.3e-09;
Matches 314; Conservative 272; Mismatches 565; Indels 528; Gaps 78;

144 FVRVYQILKINNSKRWMLPVKOS---GVPLAGTLITHCYKDLGMDPFCISLVTSV 200
804 FKQDELKVKIKTKEWVKHTSISRSQSLPLKDS---COREL-----TNLGLAP 853
201 KVPBEPSSAQLRVLLAFVASTIVSALVAEDVDNIIAKLFPYQKGLSSLPDYRAA 260
854 KI-----EMARASCSALMSQSPAD-----FVQGFPSFLGRYQAV 889
261 TYMTICQISV---KVTMENTFVNSIASQIKTKIPSLIKDGLSCLIVLQROKPSLG 317
890 -----QAVEDROOHLNELLKQCPGHAYLEFLTKLDKVLNDS-----ENKAQVSLN 935
318 KKPRLHLCNVDDLITLHGISTETVDSPLRYMLPHLVYSIIHNVTGTEGMD----- 371
936 V-----LNDLAKVERALQEKTLDEILENOXKAL-----HKLAEETKALEKNVHPD 981
372 -GOIKRHLEALITKISLKNLNDHLASLLFEEYISYSOSEEMDNKYSLNQOFLPLIR 430
982 VEKLYKQFDDVQK-----NKKLKYLVSKDLHLLEIALTL-R 1019
431 LLESKYRPTLDVLEHLKEIAD-LKQELFHQFVSLSTSGGKYQPLADSDTSLMLSLNH 489
1020 AFEND-----STVIEKMDGVKDFLMKQ-----OAAQGDADAGLQROLDQ 1058
490 PLAPRIL-----AMNHLKTKIMTKSKEGVDSFIKEVILARLGDINIDVLSAIFELF 544
1059 CSAFANELETSSSKMKKELETNLRSG-PVAGITWQTRIGL-----YOTQ 1105
545 KEHFSESEVTI---SNLNLFORA-----ELSKNGEYV----- 573
1106 LEKSKLEIATQKSRISSEOKRANLKDIAEMQEMNTQAEELYERDEPKSPPELESAN 1165
574 -EVLKIADILIKE---EILSEN-----DQLSNOVVVCLLPVYVYN----- 611
1166 EEMKAKADVLOKEVRVILKDKNIKLLAKVPSGGOELTSELNVVLENYQOLCNIRGKC 1225
612 ---DDTSAEMKIATYLSKSGICSLHPLLRGWEALENVIKSTK---PGKLIGV-----AN 661
1226 HTLEWMSCWIEHLHYD-----LETTVNLNLEEMKSTEVLPKRTDANALASL 1276
662 QKMTLLADN-----INLGDPSMLKNVEDLISVGEESFNLKQKVTFVILSVLVC--- 714
1277 ESVLHHPADNRTOIRELGTLLDGLDIIIS-EKLEAFNSRYEDLSHLAESKQSLERKQ 1335

715 CSSLKETHFFPAIRVFSLLQKKIKCESVIT-----AVEIPSEWHIELMDRGIPVE 766
1336 LQVLRETD-----QMLQVLESIGELDKQTLTYLDRIDAFQVQOE-----AQKIOAE 1383
767 LMAH---YVEELNSTR---VAEDSVFLVLSKGFITALKAPKSPKPGDIWMNEQAKED 821
1384 ISAHETLEELRRNRNQPLTSPESRTARGSQMDVLOKRLREVSTFQLFQKPAFQOR 1443
822 SRDYLHLILGL-FPMMLNGADAVFRVLMKLFIKVHLEDVFOLEKFCPSVLMYTGSSLSNP 880
1444 MLDCKRVLDGKALHLDVXDNDPDV-----IOTHLDKCKLYKLTIS-----E 1487
881 LNCVKTVLQTOALYVCAMLSQKTO-----CKHQLASISSPVVTSLLINSGPVKE--- 933
1488 VKLEVEVITIKT---GRHIVQKQOTDNPQKMDQTLSTLK-----VLVNDIGAQVTEGKQ 1537
934 -----VRRAQOCQALSGVASPFYLIIDHLISK---AEITSPDAVYVIODLATL 980
1538 DLERASQALRRMKKEAASLSBWSLATET-----ELVQKSTSGGLDLDTEISMAKNV 1590
981 FEELQREK-KLKSQKLSLTKNLLS-----CV-----YSCPSYIAKDLMKVLQGV 1025
1591 LKDLERKADLNTITESSALQNIIEGSEPIBERLCVLNAGWGRVRTWEDMCNTLMNH 1650
1026 NGENV-----LSQLLPMAEQLEKIQKEPTAVLKDEAMVHLTLQKTVESVSLNE 1077
1651 QNQEIEFDGNVAHISTWLYQAEALDDEIEKKPTS---KOEIVKRL-VSELDANLVQENV 1707
1078 DPKSLDIFIKAVHTTKELVYGMPTIQTALTEKIKPFAAISDEKVOQ---KLL---R 1129
1708 RDQALLIMNARGSSRLVE-----PKLALNRF-----EKVQHIKSAULLIAOE 1754
1130 MLFPLVNCNSHCAQTVSSVFKGISVNAQVRIELEPPDKAPLGTVOQKROKMOQK 1189
1755 PLYQCLVTEFTETFGVPFSDL-----EXTENDIENMLK-----FVEKGLTSSDEDE 1800
1190 SQDLSEVQEVGSGVWQVTLILELLOH-----KKKLRSPQILVPTLFNLLSR 1236
1801 KMDESAQ-----IEVLQGEEMLMHPMEDNKKEKRLQULLLHTYNNK--- 1845
1237 CLEPLPOGNGMEYTKOLILSCLNICOQLSPDGKIPKDLDEKRNVELIVOCIRLSE 1296
1846 -TKAIPIQORMGGLAGISSSL-----PTDYLVE---INKLILCOMDVE 1887
1297 MPQTHHALLLGTVAGIPDKVLNHTMSITFGANVMRLD---DTYSQVYINKTVKM 1352
1888 -----LSLNVPELNTAIYEDFSFO----- 1906
1353 VIPALIOSDGSSTIEVRNVEIYVKI---ISVFVDAALPHVBEHRLPLIVQVLDVTLGAE 1409
1907 -----BDSL---KNIDQDLKGBQIAVHEKQPDVILASGBEAIIQINDTLTOL 1953
1410 KFLWILLILFEQVYTVTLAAAYGERDALIEDATEMFVSCEFSYOHQIOSIMNIIQY 1469
1954 NAKH-----DRINRWYSDRKGCDFRAMEMEORQHCO-----LNDLTQW 1991
1470 LKLPEEKKEETPK-AVSFKSSQSEBMLQVFNVEHTSQDLHFKFLSYSPFSQSLSS 1527
1992 ITEBAEILLVDTCARGSIDLEKARIHQOELEV-GISSH-----QSPFALNFTG 2039
1528 NNFLKAVVSGG---PEILKGLERLLETVLGYISAVQSMER-----NADKLTVKFWRA 1579
2040 DGIYQKLSQADGSLKERTAGLNOR-----WAIYAVEVDROPRLKSGESKQWYKXVHQ 2092
1580 L-----LSKAYVLDLIDRVNALLPTETFPVIRGLV-----GNPLPSYRKALDILLNK 1626
2093 LDEITICWLTAKHAHQKRSSTELGEN-LOEILRDLTOEMEVHAEKLMNLRTELEMLSDK 2150

RESULT 3
US-09-538-092-1136
; Sequence 1136; Application US/09538092

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:52:34 ; Search time 101.893 Seconds
(without alignments)
7405.549 Million cell updates/sec

Title: US-09-603-665-5_COPY_1_1629

Sequence score: 1 MTSLAQOLQRLALPQSDASL.....NPLPSVRKALDLINKLQQ 1629

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA Main:*

- 1: /EMC_CeJerra_SIDS3/prodata/2/pubppaa/US07_PUBCOMB.pep.*
- 2: /EMC_CeJerra_SIDS3/prodata/2/pubppaa/US08_PUBCOMB.pep.*
- 3: /EMC_CeJerra_SIDS3/prodata/2/pubppaa/US09_PUBCOMB.pep.*
- 4: /EMC_CeJerra_SIDS3/prodata/2/pubppaa/US10_PUBCOMB.pep.*
- 5: /EMC_CeJerra_SIDS3/prodata/2/pubppaa/US108_PUBCOMB.pep.*
- 6: /EMC_CeJerra_SIDS3/prodata/2/pubppaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8212	100.0	2144	US-10-745-237-230	Sequence 230, App
2	8195	99.8	2144	US-11-124-368A-277	Sequence 277, App
3	7635	93.0	2044	US-11-124-368A-278	Sequence 278, App
4	7627.5	92.9	1569	US-10-450-763-45591	Sequence 45591, A
5	7266	88.5	2036	US-11-124-368A-276	Sequence 276, App
6	7266	88.5	2036	US-11-124-368A-280	Sequence 280, App
7	7266	88.5	2036	US-11-124-368A-281	Sequence 281, App
8	3694	45.0	734	US-10-108-260A-3300	Sequence 3300, App
9	2824.5	34.4	1149	US-10-128-558-167	Sequence 167, App
10	1916	23.3	897	US-10-764-425-177	Sequence 177, App
11	1603	19.5	325	US-10-128-558-358	Sequence 358, App
12	1396.5	17.0	2096	US-10-745-237-56	Sequence 36, App1
13	1396.5	17.0	2096	US-11-097-143-22518	Sequence 22518, A
14	704.5	8.6	1798	US-10-128-714-8388	Sequence 8388, App
15	704.5	8.6	1814	US-10-128-714-3388	Sequence 3388, App
16	684.5	8.3	1818	US-10-032-585-7616	Sequence 7616, App
17	674.5	8.2	2122	US-10-437-963-189782	Sequence 189782, App
18	663	8.1	135	US-10-264-049-2412	Sequence 2412, App
19	379	4.6	77	US-09-864-761-37036	Sequence 37036, A
20	337	2.9	952	US-10-424-599-248111	Sequence 248111, A
21	228	2.8	3433	US-10-408-765A-731	Sequence 731, App
22	228	2.8	3433	US-10-408-765A-732	Sequence 732, App
23	228	2.8	3433	US-10-756-149-5681	Sequence 5681, App
24	228	2.8	4131	US-10-369-493-5136	Sequence 5136, App
25	227.5	2.8	2954	US-10-650-280-1	Sequence 1, App11
26	225.5	2.7	1583	US-10-408-765A-1635	Sequence 1635, App
27	218.5	2.7	2543	US-10-828-985A-9	Sequence 9, App11

28	218	2.7	2665	US-11-124-368A-214	Sequence 214, App
29	218	2.7	2668	US-11-124-368A-215	Sequence 215, App
30	217.5	2.6	2228	US-10-511-096-2	Sequence 2, App11
31	217.5	2.6	2230	US-10-511-096-4	Sequence 4, App11
32	216	2.6	2250	US-10-511-096-6	Sequence 6, App11
33	216	2.6	2252	US-10-511-096-8	Sequence 8, App11
34	215.5	2.6	2503	US-10-828-985A-11	Sequence 11, App1
35	215	2.6	2568	US-10-828-985A-7	Sequence 7, App1
36	215	2.6	2633	US-10-450-763-36864	Sequence 36864, A
37	215	2.6	2663	US-10-723-860-749	Sequence 749, App
38	214.5	2.6	3829	US-10-408-765A-849	Sequence 849, App
39	214.5	2.6	3829	US-10-723-860-2137	Sequence 2137, App
40	213	2.6	1581	US-10-437-963-196981	Sequence 196981, A
41	212	2.6	2108	US-10-369-493-1537	Sequence 1537, App
42	210.5	2.6	2733	US-11-097-143-9003	Sequence 9003, App
43	206.5	2.5	1786	US-09-742-096-3	Sequence 3, App11
44	206.5	2.5	1786	US-11-196-400-3	Sequence 3, App11
45	206.5	2.5	1787	US-10-415-253-2	Sequence 2, App11

ALIGNMENTS

RESULT 1					
US-10-745-237-230					
Sequence 230 Application US/10745237					
Publication/No. US20050227301A1					
GENERAL INFORMATION:					
APPLICANT: Celest Limited					
APPLICANT: Glover, David					
APPLICANT: Bell, Graham					
APPLICANT: Frenz, Lisa					
APPLICANT: Midgley, Carol					
TITLE OF INVENTION: Cell Cycle Progression Proteins					
FILE REFERENCE: P0158190 CYK					
CURRENT APPLICATION NUMBER: US/10/745,237					
CURRENT FILING DATE: 2003-12-23					
PRIOR APPLICATION NUMBER: US 60/439,123					
PRIOR FILING DATE: 2003-01-10					
PRIOR APPLICATION NUMBER: US 60/468,402					
PRIOR FILING DATE: 2003-05-06					
NUMBER OF SEQ ID NOS: 600					
SOFTWARE: Patent version 3.1					
SEQ ID NO 230					
LENGTH: 2144					
TYPE: PRT					
ORGANISM: Homo sapiens					
FEATURE:					
OTHER INFORMATION: Q9H583					
US-10-745-237-230					
Query Match					
Best Local Similarity 100.0%; Score 8212; DB 5; Length 2144;					
Matches 1629; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MTSLAOLQRLALPQSDASLSRDEVASLFDPKKATIRDTAFAGTGTLEELGIDP	60		
DB	1	MTSLAOLQRLALPQSDASLSRDEVASLFDPKKATIRDTAFAGTGTLEELGIDP	60		
QY	61	SFQGFAPLPSQAKTLERSVQTKAVNKQDENISFLHLSPYFLKPAOKCLEWLIHR	120		
DB	61	SFQGFAPLPSQAKTLERSVQTKAVNKQDENISFLHLSPYFLKPAOKCLEWLIHR	120		
QY	121	FHHLNQSGLICVLPYHETRIFFAVIQLKTKNNGHRFWFLLPVQSGVPLAKGTLIT	180		
DB	121	FHHLNQSGLICVLPYHETRIFFAVIQLKTKNNGHRFWFLLPVQSGVPLAKGTLIT	180		
QY	121	FHHLNQSGLICVLPYHETRIFFAVIQLKTKNNGHRFWFLLPVQSGVPLAKGTLIT	180		
DB	121	FHHLNQSGLICVLPYHETRIFFAVIQLKTKNNGHRFWFLLPVQSGVPLAKGTLIT	180		
QY	181	HCKKDGFMDPISLTVKSVKFAEYPGSSAQRLVLAFASTIVSALVAEDVSDNIIA	240		
DB	181	HCKKDGFMDPISLTVKSVKFAEYPGSSAQRLVLAFASTIVSALVAEDVSDNIIA	240		
QY	241	KLPFYIOKGLSKSLPYRAATYMIICQISVKVTMENTFVNSLASQIITKLTJPSLIXDG	300		
DB	241	KLPFYIOKGLSKSLPYRAATYMIICQISVKVTMENTFVNSLASQIITKLTJPSLIXDG	300		

```

Db 241 KLFPYIQKGLSSLPDYRAATYMIICQISVKYMENTFVNSLASQIITKTKIPSLIKDG 300
Qy 301 LSCLIIVLQROKPESELGKKPFPHLGNVPDLITLHGISETYDVSPLLRYMLPHLVASIIH 360
Db 301 LSCLIIVLQROKPESELGKKPFPHLGNVPDLITLHGISETYDVSPLLRYMLPHLVASIIH 360
Qy 361 HVTGEETGMDGOQIYKRLHAILTKISLKNNDHLLASLLEBYISYSSQOEEMDSNKVSL 420
Db 361 HVTGEETGMDGOQIYKRLHAILTKISLKNNDHLLASLLEBYISYSSQOEEMDSNKVSL 420
Qy 421 LNEOGLPLIRLESKXPRTLDVLYLEHILKEIADLKKOELEHQVSLSTSGGKYGOGLAOSD 480
Db 421 LNEOGLPLIRLESKXPRTLDVLYLEHILKEIADLKKOELEHQVSLSTSGGKYGOGLAOSD 480
Qy 481 TSLMLSLNHPILAPVAILAMNHLKKIMKTSKEGVDESFIKEAVLARGDNDIVLSAISA 540
Db 481 TSLMLSLNHPILAPVAILAMNHLKKIMKTSKEGVDESFIKEAVLARGDNDIVLSAISA 540
Qy 541 PEIFKEHFSSEYTTISNLLNFORAELSKNGEWEVYLKIAADLILKEIILSENDOLSNQV 600
Db 541 PEIFKEHFSSEYTTISNLLNFORAELSKNGEWEVYLKIAADLILKEIILSENDOLSNQV 600
Qy 601 VCLLPFVIVINDDPESAEMKIAIYLSKSGICSLHPLLGMEBALENVIKSTKPGKLGVA 660
Db 601 VCLLPFVIVINDDPESAEMKIAIYLSKSGICSLHPLLGMEBALENVIKSTKPGKLGVA 660
Qy 661 NQKMIELADNINLGDPSMLKMEVDLISVGEESFNLKOKVTPHVLISVLVSCSSLIKE 720
Db 661 NQKMIELADNINLGDPSMLKMEVDLISVGEESFNLKOKVTPHVLISVLVSCSSLIKE 720
Qy 721 THPPAIRVSELLOKKIKKLESVITAVEIPSEMHIELMDRGIPEVLAHYVEELNSTOR 780
Db 721 THPPAIRVSELLOKKIKKLESVITAVEIPSEMHIELMDRGIPEVLAHYVEELNSTOR 780
Qy 781 VAVDSVFLVLSLKKFIYALKAPKSPFGDITMWNNEQEKEDRDYLLHILIGLFEMMLNGA 840
Db 781 VAVDSVFLVLSLKKFIYALKAPKSPFGDITMWNNEQEKEDRDYLLHILIGLFEMMLNGA 840
Qy 841 DAVHFRVLMKLFIKVHLEDFOLFKEGCVLMTYSSLSNPLNCSVKTVLQTOALVYGAM 900
Db 841 DAVHFRVLMKLFIKVHLEDFOLFKEGCVLMTYSSLSNPLNCSVKTVLQTOALVYGAM 900
Qy 901 LSSQKTOCKHOLASSPVSTSLINLGSVKEVRRALIOCCIALSGVASPYYLIIHILI 960
Db 901 LSSQKTOCKHOLASSPVSTSLINLGSVKEVRRALIOCCIALSGVASPYYLIIHILI 960
Qy 961 SKAEBITSDAAYVIOIDLTLFEELOREKKLKSQKLESETLKULSCVYSCPSYIAKDLMK 1020
Db 961 SKAEBITSDAAYVIOIDLTLFEELOREKKLKSQKLESETLKULSCVYSCPSYIAKDLMK 1020
Qy 1021 VLOGVNGEMVLSQLPMAEQULLEKIQKEPTAVLKDAMVLAHLTLGKYNFVSLLNEDPK 1080
Db 1021 VLOGVNGEMVLSQLPMAEQULLEKIQKEPTAVLKDAMVLAHLTLGKYNFVSLLNEDPK 1080
Qy 1081 SLDFIKAVHTTKELIYAGMPTQITALEKITKPPFAAISDEKVQOKLRLMFDLLVNCN 1140
Db 1081 SLDFIKAVHTTKELIYAGMPTQITALEKITKPPFAAISDEKVQOKLRLMFDLLVNCN 1140
Qy 1141 SHCAQTVSVFEGISVNAEQVRIELEPPDKAKPLGTVOQKRQKQKQKKSODLESVOEGV 1200
Db 1141 SHCAQTVSVFEGISVNAEQVRIELEPPDKAKPLGTVOQKRQKQKQKKSODLESVOEGV 1200
Qy 1201 GSYWQVTLILLELOHKKKLSPOQLVPTLPMILSRCEPLRPOEOGNNBYTQOLITLSCIL 1260
Db 1201 GSYWQVTLILLELOHKKKLSPOQLVPTLPMILSRCEPLRPOEOGNNBYTQOLITLSCIL 1260
Qy 1261 NICQGLSDGGKIPKDIIDEEKFNVELIYQICIRLSEMOTHHALLLGTAVAGIPDPKVL 1320
Db 1261 NICQGLSDGGKIPKDIIDEEKFNVELIYQICIRLSEMOTHHALLLGTAVAGIPDPKVL 1320
Qy 1321 HNINSIPTFMGANVVRDLDTYSFOVINKTVKXVIBALIQSDSGDSIEVSRNVEEIVKXI 1380
Db 1321 HNINSIPTFMGANVVRDLDTYSFOVINKTVKXVIBALIQSDSGDSIEVSRNVEEIVKXI 1380

```

```

Qy 1381 SVFVDALPHVBEHRRLPILVOLVDTLGAKEFLMILLILPEQVYTKTVLAAAYGEKDAL 1440
Db 1381 SVFVDALPHVBEHRRLPILVOLVDTLGAKEFLMILLILPEQVYTKTVLAAAYGEKDAL 1440
Qy 1441 EADTFEWFVSVCCESSVOHQIOSLNNIIQYLLKLPBEKEETIPKAVSPKSSQOEMLQVF 1500
Db 1441 EADTFEWFVSVCCESSVOHQIOSLNNIIQYLLKLPBEKEETIPKAVSPKSSQOEMLQVF 1500
Qy 1501 NVEHTSKOLHFPFLSVSFMQSLSSNNPLKTVVSGGPEIILGLBERLLETVLGYISA 1560
Db 1501 NVEHTSKOLHFPFLSVSFMQSLSSNNPLKTVVSGGPEIILGLBERLLETVLGYISA 1560
Qy 1561 VAQSMERNADLTVKFWRALISKAAYDLDRYNALLPTEFTIPVIRGLVGNPLPSVRKAL 1620
Db 1561 VAQSMERNADLTVKFWRALISKAAYDLDRYNALLPTEFTIPVIRGLVGNPLPSVRKAL 1620
Qy 1621 DLNNKLOQ 1629
Db 1621 DLNNKLOQ 1629

RESULT 2
US-11-124-368A-277
; Sequence 277, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Lake
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE REFERENCE: C0001524
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/566,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 277
; LENGTH: 2144
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-368A-277

Query Match 99.8%; Score 8195; DB 6; Length 2144;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1626; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MTSIAQOLQRLAPQSDASLISRDEVALLPDPKRAATIDRTAFAGCTGLBEILGIDP 60
Db 1 MTSIAQOLQRLAPQSDASLISRDEVALLPDPKRAATIDRTAFAGCTGLBEILGIDP 60
Qy 61 SFEQPEAPLPSQAKTIERSVQTVKAVNKQDENTSLFIHLSPEFLKPAQKLEMLIHR 120
Db 61 SFEQPEAPLPSQAKTIERSVQTVKAVNKQDENTSLFIHLSPEFLKPAQKLEMLIHR 120
Qy 121 FHIHLYNODSLIACVLPEYHETRIFVRVYQILKINNSKRWFWMLPVYQSGVPLAKGTLIT 180
Db 121 FHIHLYNODSLIACVLPEYHETRIFVRVYQILKINNSKRWFWMLPVYQSGVPLAKGTLIT 180
Qy 181 HCYDGLGFMDFICGLVTKSVKVPFAEYRGSSAQQLAVLAFYASTIVSLVAEEDVSDNIIA 240
Db 181 HCYDGLGFMDFICGLVTKSVKVPFAEYRGSSAQQLAVLAFYASTIVSLVAEEDVSDNIIA 240
Qy 241 KLFPYIQKGLSSLPDYRAATYMIICQISVKYMENTFVNSLASQIITKTKIPSLIKDG 300
Db 241 KLFPYIQKGLSSLPDYRAATYMIICQISVKYMENTFVNSLASQIITKTKIPSLIKDG 300
Qy 301 LSCLIIVLQROKPESELGKKPFPHLGNVPDLITLHGISETYDVSPLLRYMLPHLVASIIH 360
Db 301 LSCLIIVLQROKPESELGKKPFPHLGNVPDLITLHGISETYDVSPLLRYMLPHLVASIIH 360

```

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:53:49 ; Search time 9.06679 Seconds
(without alignments)
4243.670 Million cell updates/sec

Title: US-09-603-665-5_COPY_1_1629

Perfect score: 8212

Sequence: 1 MTSIAQOLQRLALPQSDASL.....NPLPSVRKALDLNNKQ 1629

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 103426 seqs, 23619683 residues

Total number of hits satisfying chosen parameters: 103426

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications_AA_New*

- 1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 7: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_NEW_PUB pep.*
- 8: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3694	45.0	734	7 US-11-293-697-3300	Sequence 3300, App
2	206.5	2.5	3113	6 US-10-505-928-325	Sequence 325, App
3	188.5	2.3	2209	7 US-11-301-554-1903	Sequence 1903, App
4	171	2.1	1325	6 US-10-449-902-51794	Sequence 51794, A
5	166.5	2.0	1206	6 US-10-449-902-43305	Sequence 43305, A
6	165	2.0	1176	6 US-10-449-902-46322	Sequence 46322, A
7	163.5	2.0	1328	6 US-10-504-973-32	Sequence 32, Appl
8	159.5	1.9	1116	6 US-10-505-928-784	Sequence 784, App
9	158.5	1.9	1116	6 US-10-449-902-47106	Sequence 47106, A
10	157.5	1.9	980	7 US-11-293-697-3967	Sequence 2967, App
11	157	1.9	1205	7 US-11-293-697-3967	Sequence 2967, App
12	155	1.9	764	6 US-10-449-902-45961	Sequence 45961, A
13	155	1.9	1842	6 US-10-511-937-2829	Sequence 2829, App
14	152.5	1.9	1230	6 US-11-289-102-378	Sequence 378, App
15	151.5	1.8	1018	6 US-10-449-902-47146	Sequence 47146, A
16	150	1.8	1250	6 US-10-449-902-43118	Sequence 43118, A
17	147.5	1.8	1443	6 US-10-471-571A-4624	Sequence 14, Appl
18	146.5	1.8	1146	6 US-10-486-020-14	Sequence 14, Appl
19	141.5	1.7	1056	6 US-10-449-902-45516	Sequence 45516, A
20	141	1.7	1346	6 US-10-449-902-55021	Sequence 55021, A
21	140.5	1.7	1663	6 US-10-505-928-227	Sequence 227, App
22	139.5	1.7	1174	6 US-10-449-902-49996	Sequence 49996, A
23	139.5	1.7	1346	6 US-10-449-902-55428	Sequence 55428, A
24	139.5	1.7	2364	7 US-11-289-102-242	Sequence 242, App
25	136.5	1.7	1575	6 US-10-505-928-257	Sequence 257, App

26	136	1.7	1315	6 US-10-527-411-141	Sequence 141, App
27	135.5	1.7	838	6 US-10-449-902-45252	Sequence 45252, A
28	135.5	1.7	1479	6 US-10-471-571A-2392	Sequence 2392, App
29	134.5	1.6	763	6 US-10-449-902-45594	Sequence 45594, A
30	134	1.6	1166	6 US-10-449-902-37528	Sequence 37528, A
31	133.5	1.6	712	7 US-11-293-697-3282	Sequence 3282, App
32	133.5	1.6	1085	6 US-10-505-928-343	Sequence 343, App
33	133.5	1.6	1158	6 US-10-471-571A-4944	Sequence 4944, App
34	132.5	1.6	784	6 US-10-449-902-43318	Sequence 43318, A
35	132.5	1.6	926	6 US-10-449-902-43257	Sequence 43257, A
36	132	1.6	889	6 US-10-449-902-36405	Sequence 36405, A
37	131.5	1.6	836	7 US-11-105-233-154	Sequence 154, App
38	131.5	1.6	995	6 US-10-471-571A-3910	Sequence 3910, App
39	131.5	1.6	1866	6 US-10-511-937-2968	Sequence 2968, App
40	130.5	1.6	726	7 US-11-334-351-2	Sequence 2, Appl1
41	130.5	1.6	3176	7 US-11-105-233-155	Sequence 155, App
42	130	1.6	732	7 US-11-203-828-7	Sequence 7, Appl1
43	130	1.6	732	7 US-11-221-332-102	Sequence 102, App
44	130	1.6	882	7 US-11-122-986-140	Sequence 140, App
45	130	1.6	882	7 US-11-122-986-142	Sequence 142, App

ALIGNMENTS

RESULT 1
US-11-293-697-3300
; Sequence 3300, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3300
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-293-697-3300

Query Match 45.0%; Score 3694; DB 7; Length 734;
Best Local Similarity 100.0%; Pred. No. 1.7e-208;
Matches 734; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTSLAQQOLQRLALPQSDASLSRDEVASLFDPEKATIDRDTAFAGTGLEELCIDP	60
DB	1	MTSLAQQOLQRLALPQSDASLSRDEVASLFDPEKATIDRDTAFAGTGLEELCIDP	60
QY	61	SFQGFAPLPSQAKTLERSVQTKAVNKQIDENISLHLSPFKLPQCKLEMLIHR	120
DB	61	SFQGFAPLPSQAKTLERSVQTKAVNKQIDENISLHLSPFKLPQCKLEMLIHR	120
QY	121	FHHLYNODSLIACVLPYHETRIPIVRYIOLKINNSGHRFWMLPVQSGVPLAKGLTIT	180
DB	121	FHHLYNODSLIACVLPYHETRIPIVRYIOLKINNSGHRFWMLPVQSGVPLAKGLTIT	180
QY	181	HCVYDLGFMDFICSLVTKSVKFAEPYGGSAQRLVLLAFVASTIVSALVAEDVSDNIIA	240
DB	181	HCVYDLGFMDFICSLVTKSVKFAEPYGGSAQRLVLLAFVASTIVSALVAEDVSDNIIA	240
QY	241	KLPFYIOGKGLSSLPDYRAATYMTICQISVAVTMENTFVNSLASQIITLTKRISLIDG	300
DB	241	KLPFYIOGKGLSSLPDYRAATYMTICQISVAVTMENTFVNSLASQIITLTKRISLIDG	300
QY	301	LSCLIVLQROKPSLGKPPHLCNVPDLITLHGISETVDSVSLRLYMLPHLVVSIH	360
DB	301	LSCLIVLQROKPSLGKPPHLCNVPDLITLHGISETVDSVSLRLYMLPHLVVSIH	360

Qy	361	HTGGEETECMDQIKRHHALITTSIKNNLDHLASILEEYITSVSSQEMDSNVSL	420
Db	361	HYTGESTECMDQIKRHLREALITYISLKNNDLHLASILLEEYISVSSQEMDSNVSL	420
Qy	421	LNBOFLPLIRLLESKYPRTLDVLEEHKEIAIDLKKOELPHQFVSLSTSGCKYQFLADSD	480
Db	421	LNBOFLPLIRLLESKYPRTLDVLEEHKEIADLKKOELPHQFVSLSTSGCKYQFLADSD	480
Qy	481	TSLSMSLNHPLAPVILAMNHLLKTIKTSKSGVDSEFIKEAVLARLDGDNIDVVLSAISA	540
Db	481	TSLSMSLNHPLAPVILAMNHLLKTIKTSKSGVDSEFIKEAVLARLDGDNIDVVLSAISA	540
Qy	541	FELFKEHSESEYITINLNLFORAELSKNGEYEVLTAAIDILKEEILISENQLSNQVV	600
Db	541	FELFKEHSESEYITINLNLFORAELSKNGEYEVLTAAIDILKEEILISENQLSNQVV	600
Qy	601	VCLLPFVVINNDTESAEKIAIAYLSKSGICSLPHLLRGMEALENVIKSTKCGKLIGVA	660
Db	601	VCLLPFVVINNDTESAEKIAIAYLSKSGICSLPHLLRGMEALENVIKSTKCGKLIGVA	660
Qy	661	NOKMTELLADNINIGDPSSMLKMVEDLISVGEESFNLKOKVTFPHVILSVLVSCSSLIKE	720
Db	661	NOKMTELLADNINIGDPSSMLKMVEDLISVGEESFNLKOKVTFPHVILSVLVSCSSLIKE	720
Qy	721	THFPFAIRVFSLLQ 734	
Db	721	THFPFAIRVFSLLQ 734	

RESULT 2
US-10-505-928-325
Sequence 325, Application US/10505928

```

?
?
? APPLICANT: Ludwig Institute for Cancer Research et al
? TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
? FILE REFERENCE: 28967/39178
? CURRENT APPLICATION NUMBER: US/10/505,928
? CURRENT FILING DATE: 2004-08-27
? PRIOR APPLICATION NUMBER: US 60/363,019
? PRIOR FILING DATE: 2002-03-07
? NUMBER OF SEQ ID NOS: 866
? SOFTWARE: PatentIn 3.2
? SEQ ID NO 325
? LENGTH: 3113
? TYPE: PRT
? ORGANISM: Homo sapiens
? OS-10-505-928-325

```

Query Match	2.5% ; Score 206.5 ; DB 6 ; Length 3113 ;
Best Local Similarity	18.4% ; Pred. No. 0.0016 ;
Matches 309 ;	Conservative 294 ; Mismatches 584 ; Indels 489 ; Gaps 75 ;
QY	266 CQISVKVMTMENTFVNSLASQIIKTLTKIPSLIKDGLSCGLIVLQROKESLGGKPPHLC 325
Db	1425 CQSSKSKSELOTYDVSLSKENTVLSTLNRNFGD--LVEMOLGLGEGL----- 1474
QY	326 NVVDLITILHGISTYDVSPLLRNMLPRLVVISIHNVTGEETEGDQGYRHRLEALITK 385
Db	1472 -----VSLSSCYVPDSSLSLGL-DSSFYALLFQ-TGD 1504
QY	386 ISLKNNT-----DHLASLFEETSYSS----- 409
Db	1505 MSLSLNIEGAVSANQSCVDVEVFCSSLOEBENTTRKRETPSPAPKAGVEELESLECVYRQSLK 1564
QY	410 -OEBMSNKVSLINEOFLPIRLIESKPRPLDVVLEELKELADLKQGLFHOVSLST 468
Db	1565 LEEKMEQGI-MKKKEIQELFQLSLSE-RQSLDLCKRQYLS-----NEOMQKLTSTV 1616
QY	469 SGGKYQFLADSDTSLMSLNLHPLAPVRLIANNHLKTKMTKSEGVDSFIEAVLARIQD 528
Db	1617 LEMESKLAAEKKQTEQSLLELVARLO-----OGDLS-----SRSLGI 1657

Qy	529	DNDIVLVSALSAPEIIFREHFSSEYT-----ISNLNLMPORABLSKNGEVYLKLAAD	561
Db	1658	DTBEAIGRNESCDSIKKH-TSETTEBTPKHDVQICDQAQOOLN-----LDIEKTET	1711
Qy	582	ILIKE--EILSENOLSNQVAVCLLPVIVINNDPTEAEMKIAIYLSGSCISLHPL-IR	638
Db	1712	GALKPTECGSGEOPDIN-----YEPGEDTIGSSSECS-ELFSFGNALVPMDFL	1762
Qy	639	GWEBALENV--IKSYPKGL-----IGVANQKITEIILAD-----NINLGDPSMLKM-	683
Db	1763	GNQEDIHNLQURVETSENLRILHVEDRDKYESLNNKEKELDSKYLHQEOVLMTKIE	1822
Qy	684	-----VEBLLISGEBESFNLKOKYF-----HYILSVLSCSSSLK-----ETHPFAIRV---	729
Db	1823	ACIELEKTIYVGLKXKNSDLSSEKLEYFSCDHOELLQRYVETSEGLNSDLBEMADKSRDIEG	1882
Qy	730	-----FSLLOKKIKKLESYITVAEIPSEWHIEIEMDRGIPVELMAHYVEELN	776
Db	1883	DNYAKVNDMSMKERLDVENEELSRRSRSEKASIE-----HEALYIEADL-----EVV	1927
Qy	777	STORVAV-----EDSVELVFSGLKFIYALKAPKSPKGD-----WNNPEOLKSDSDYL	826
Db	1928	QTEKLCLEKONENKQKIVICLIEBELSVITSEBRNDRGELPTMSKKTALDOLSEMKEXT	1987
Qy	827	HLILGLEMMLNGADAVHFRVIMLFLFKYHLEDFVQLFKFCVSVMVTGSSLSNPLNCSVK	886
Db	1988	Q-----ELSSHQSECHCIQVNEAEVYEKTE-----LLQTTSSDVSELLKDKTH	2031

Dh	2032	IQEHLTQLEKSDQALSTKTCOLENQIOLNKKKEKLLVKESESLQARLSESDPYEKLTKNSA	20911
Oy	919	VWTSLL-----INSGPVKEVR--RAIQCLQALSGVAPFYLIIDHLISKAEIISDA	970
Dh	2092	LEALVKEGFARLSTQEBEVQOLRGIEKLRIADEKQOL--HIEKLEKEREBEN	2148
Oy	971	AYVQIADATLFEELQREKKL-KSHQKL-----SETLKNLSCVYSCPSYIANKL	1018
Dh	2149	----DSLKQKVNERLELQMSSENOELVILDAENSKAVEFLTKQIEMARSLKVFEJDL	2204
Oy	1019	M-----KVLQGVNGEM-VLSOLLPMABQLELKIQKEPTAV-LKDEAM-----VJHL	10623
Dh	2205	VTLRSEKENTTKQIQEKQOGSELDPKLLSFFSKLLE--EKQBQAEIQJKEBSTAVEMQON	22623
Oy	1063	TLKGYNFVSILNED-----PKSLDIFIKAVTTKELYAGMPTQITALEKITTKBPFA	1116
Dh	2263	QKELNNAVAALCGDQEIIMATDSQSLDPIIEBHQRL-----NSIEKURLALEA	2311
Oy	1117	AISDEKVOQKLKRLFPDLVNCNKSHCQIVSVYFKIS-----VWAEQVRIELEPRD	1165
Dh	2312	----DEKCOLCVLQOLAKE-----SEHNADLLKQGVNELERLEIARTNOEHALAEANSK	23623
Oy	1170	-----KAKPLGTVQQRQRMQOOKSQDLESVQVNGSWORVTLLLELLOHKKLRSP	1223
Dh	2363	GEVETLKAKIEGHTQSLRGLEL-----DVVTIRS-----EKNLNLNELQKEBERISEL	2410
Oy	1224	QILVPTLFNLLSRCELEPQEOGNMEYTKOLISCLINCOKUSPDGSKI PKD--ILDEE	1281
Dh	2411	EIINSSPENT-----LOEKEQEKYQOMKESSTAMEMLOTOKELNERVAALHNDEACKAK	2466
Oy	1282	KPNVELLVOCIRLSE-----MPTNTHALLLLGTVAQIF-----PDVLTNII	1323
Dh	2467	KONISSQVECELEBKALLOGLDBEAKNNYIVLOSANGLLQVQVEDGQKLEKQDEBISRL	2526
Oy	1324	MSIIFTEGANVMRLDPT-----YSFOVINKTVAMVP-ALIQSDSG--DSIEY-	1368
Dh	2527	KNQIQDOEOLVSKISQVEGEHOLWKEONLEIRNLVTALBEKQIQVLOSKNASLDOTLEVLQ	2586
Oy	1369	--SRNVF--EIVWKISVVDLPHNPE-----HRRPLFVLQVDTPLAQEK--FL	1412
Dh	2587	SSYKCNLENLELTMDKMSFEVKVNNKTAKETELQREMHMAQCTABLQELHSECKRILA	2646
Oy	1413	WILLIILFEQYVTKVLAAYGEXDALILEADTEBFSVCC-----	1452

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:41:09 ; Search time 144.903 Seconds
(without alignments)

6765.010 Million cell updates/sec

Title: US-09-603-665-5
10803

Sequence: 1 MTSLAQQLQRLALPQSDASL.....CQKTIQQLKLVLPGLPSYF 2144

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq.8:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*
10: geneseqp2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10799	99.9	2144	4	AAB85029
2	10799	99.9	2144	8	ADQ89800 Protein e
3	10791	99.9	2144	8	AEEO2859 Antagonis
4	10782	99.8	2144	8	ADSD34902 Human S-1
5	10222	94.6	2044	8	ADSD34907 Human aut
6	9853	91.2	2036	8	ADSD34905 Human aut
7	9853	91.2	2036	8	ADSD34903 Human aut
8	9853	91.2	2036	8	ADSD34904 Human aut
9	7756.5	71.8	1569	4	ABG15232 Novel hum
10	6124.5	56.7	1229	8	ADQ66400 Novel hum
11	5026.5	46.5	1149	8	ADQ66400 Novel hum
12	5026.5	46.5	1149	8	ADQ66400 Novel hum
13	4506	41.7	897	8	ADQ66220 T cell ac
14	4488	41.6	887	8	ADQ66218 T cell ac
15	4494	41.6	897	8	ADQ66152 T cell ac
16	3694	34.2	734	7	ADMO64615 Human pro
17	3694	34.2	734	9	AEEO2884 Human CDN
18	3102	28.7	624	9	AEEO2884 Protein t
19	2585	23.9	515	8	AAW54099 Homo sapi
20	2061	19.1	408	8	ADSD34906 Human aut
21	2051.5	19.0	206	4	ABBS5242 Drosophi
22	2051.5	19.0	206	8	ADQ89606 Antagonis
23	1779	16.5	349	8	ADP24166 PRO polyp

24	1778	16.5	349	4	AAB92729	Aab92729 Human pro
25	1603	14.8	325	7	ADE09054	Ade09054 Novel pro
26	1603	14.8	325	9	ADU40573	Adu40573 Novel hum
27	1086	10.1	1798	6	ABJ26330	Abj26330 Aspergill
28	1085	10.0	1814	6	ABJ25730	Abj25730 Aspergill
29	1056.5	9.8	1818	5	ABP73779	Abp73779 Candida a
30	956	8.8	1769	6	ABR52974	Abp73779 Candida a
31	956	8.8	1769	7	ADK62658	Adk62658 Disease t
32	663	6.1	135	5	ADP41280	Adp41280 Human ova
33	434	4.0	437	8	ADX75081	Adx75081 Plant ful
34	379	3.5	77	4	AAW17368	Aaw17368 Peptide #
35	379	3.5	77	4	ABR36384	Abb36384 Peptide #
36	379	3.5	77	4	AAW29882	Aam29882 Peptide #
37	379	3.5	77	4	ABR31185	Abb31185 Peptide #
38	379	3.5	77	4	ABR21738	Abb21738 Protein #
39	379	3.5	77	4	AAW69542	Aam69542 Human bon
40	379	3.5	77	4	AAW57146	Aam57146 Human bra
41	379	3.5	77	4	ABG51215	Abg51215 Human liv
42	379	3.5	77	4	AAW05057	Aam05057 Peptide #
43	379	3.5	77	5	ABG39167	Abg39167 Human pep
44	309.5	2.9	2228	7	ABR61599	Abp61599 Human gol
45	309.5	2.9	2230	6	ABU07445	Abu07445 Protein d

ALIGNMENTS

RESULT 1
ID AAB85029 standard; protein; 2144 AA.
AC AAB85029;
DT 06-AUG-2001 (first entry)
XX
XX Protein encoded by BAP28 cDNA consisting of exons 1 to 45.
XX BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FH Misc-difference 1694 /label= Ser or Asn
FT Misc-difference 1654 /label= Ala or Val
FT Misc-difference 1967 /label= Asp or Asn
FT Misc-difference 2017 /label= Gly or Glu
FT
XX
XX WO200100669-A2.
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-IB001183.
XX
XX 25-JUN-1999; 99US-0141323P.
XX 18-JAN-2000; 2000US-0176880P.
XX
XX (GEST) GENSET.
XX Barry C, Bouguetieret L, Chumakov I, Cohen-Akenine A;
XX MPI; 2001-367032/38.
XX N-PSDB; AAF83909, AAF83910.
XX
XX New BAP28 polynucleotides and polypeptides overexpressed in prostate
XX cancer cells for diagnosing prostate tumors, e.g. by hybridization or
XX polymerase chain reaction assays.
XX Claim 14; Page 297-304; 349pp; English.

CC The invention is directed to BAP28 polypeptides, BAP28 polynucleotide sequences and regulatory region located at the 3' and 5' ends of the BAP28 coding region. The BAP28 polypeptides can be expressed by standard recombinant methodology. BAP28 polynucleotides and polypeptides have been found to be over expressed in prostate tumour cells, therefore levels of BAP28 expression and/or activity may be assayed (e.g. by polymerase chain reaction (PCR)) to diagnose patient suffering from or susceptible to prostate cancer. Antibodies specific for the BAP28 polypeptides are useful as diagnostic reagents. Biallelic markers of the BAP28 gene are encoded by a first CDNA sequence of the BAP28 gene consisting of the exons 1 to 45

XX Sequence 2144 AA:

Query Match 99.9%; Score 10799; DB 4; Length 2144;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSIAQOQIQRALPQSDASLSRDEVASLFPDPEKATIDRDTAFATIGCTGLEELGIDP 60
DB 1 MTSIAQOQIQRALPQSDASLSRDEVASLFPDPEKATIDRDTAFATIGCTGLEELGIDP 60
QY 61 SFEQFEAPLFSQAKTLERSVQTKAVNKQDBENISFLIHSPYFLKPAQKCEMLIHR 120
DB 61 SFEQFEAPLFSQAKTLERSVQTKAVNKQDBENISFLIHSPYFLKPAQKCEMLIHR 120
QY 121 FPHILYNODSLIACVLPHEHTRIFVRVIOQLKINNSKRMFMILPVKOSGVPILAKGTLIT 180
DB 121 FPHILYNODSLIACVLPHEHTRIFVRVIOQLKINNSKRMFMILPVKOSGVPILAKGTLIT 180
QY 181 HCYKDIQFMDFICSLVTKSVKFAEPGSSAQLRVLLAFVASTIVSALVAEDVSDNIIA 240
DB 181 HCYKDIQFMDFICSLVTKSVKFAEPGSSAQLRVLLAFVASTIVSALVAEDVSDNIIA 240
QY 181 HCYKDIQFMDFICSLVTKSVKFAEPGSSAQLRVLLAFVASTIVSALVAEDVSDNIIA 240
DB 181 HCYKDIQFMDFICSLVTKSVKFAEPGSSAQLRVLLAFVASTIVSALVAEDVSDNIIA 240
QY 241 KLFPYIOQKGLKSSLPDRAATYMIICQISVKTMENTFVNSLAQIILTKIKISILKDG 300
DB 241 KLFPYIOQKGLKSSLPDRAATYMIICQISVKTMENTFVNSLAQIILTKIKISILKDG 300
QY 241 KLFPYIOQKGLKSSLPDRAATYMIICQISVKTMENTFVNSLAQIILTKIKISILKDG 300
DB 241 KLFPYIOQKGLKSSLPDRAATYMIICQISVKTMENTFVNSLAQIILTKIKISILKDG 300
QY 301 LSCILVILORQPSLSGKPPPHLCNVPLITIIHGISSETVDSPILRMYLPHLVSTIHH 360
DB 301 LSCILVILORQPSLSGKPPPHLCNVPLITIIHGISSETVDSPILRMYLPHLVSTIHH 360
QY 301 LSCILVILORQPSLSGKPPPHLCNVPLITIIHGISSETVDSPILRMYLPHLVSTIHH 360
DB 301 LSCILVILORQPSLSGKPPPHLCNVPLITIIHGISSETVDSPILRMYLPHLVSTIHH 360
QY 361 HVTGEETEGMDGQIYKRLHLEILTKISLKNLDDLHLLASLFEETIYSSSGQEMDSNKYSL 420
DB 361 HVTGEETEGMDGQIYKRLHLEILTKISLKNLDDLHLLASLFEETIYSSSGQEMDSNKYSL 420
QY 421 LNEQFLILIRLESKYPRTLDPVLEEHKETAIDLKQELPHQFVSLSTSGKYOFLADSD 480
DB 421 LNEQFLILIRLESKYPRTLDPVLEEHKETAIDLKQELPHQFVSLSTSGKYOFLADSD 480
QY 481 TSMLSLNHPRLAPVRIILAMNHLKKIMKTSKEGVDSFIKEAVLARLDGDDIDVYLSAISA 540
DB 481 TSMLSLNHPRLAPVRIILAMNHLKKIMKTSKEGVDSFIKEAVLARLDGDDIDVYLSAISA 540
QY 541 FEITKEHPSSEVTISNLNLFQRAELSKNGEYEVLLKADIIILKEELISNDOLSNQV 600
DB 541 FEITKEHPSSEVTISNLNLFQRAELSKNGEYEVLLKADIIILKEELISNDOLSNQV 600
QY 601 VCLLPFVAVINDDPESAMKIAIYLSKSGICSLPHLKGMEALENVIKSTKPKGLIGVA 660
DB 601 VCLLPFVAVINDDPESAMKIAIYLSKSGICSLPHLKGMEALENVIKSTKPKGLIGVA 660
QY 661 NOKMIETLLADNINIGDPSMLKMWEDLISVGEESFNLKQKVTFHVILISLVSCSSSLKE 720
DB 661 NOKMIETLLADNINIGDPSMLKMWEDLISVGEESFNLKQKVTFHVILISLVSCSSSLKE 720
QY 721 THFPFARVPSLLQKIKKLESVITTAIVEIPEEMHILMDRGIVELMAHVEELNSQR 780
DB 721 THFPFARVPSLLQKIKKLESVITTAIVEIPEEMHILMDRGIVELMAHVEELNSQR 780
QY 781 VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQKEDSRDYLLHLILGFEMMLNGA 840
DB 781 VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQKEDSRDYLLHLILGFEMMLNGA 840

DB 781 VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQKEDSRDYLLHLILGFEMMLNGA 840
QY 841 DAHFVRVLMKLFIKVHLEDVFOLEFKCSVMTYTGSSLSNPLNGSVKTVLOTQALVYGCAM 900
DB 841 DAHFVRVLMKLFIKVHLEDVFOLEFKCSVMTYTGSSLSNPLNGSVKTVLOTQALVYGCAM 900
QY 901 LSSQKTQCKQLASISSPVVTSLLINSGSVKEVRRRAIOTQALSGVASPFYIIDLHI 960
DB 901 LSSQKTQCKQLASISSPVVTSLLINSGSVKEVRRRAIOTQALSGVASPFYIIDLHI 960
QY 961 SKAEETSDAAVYIOTDALTPEELOREKUKSHOKSETLKNILSCYSCPSYAKDLMK 1020
DB 961 SKAEETSDAAVYIOTDALTPEELOREKUKSHOKSETLKNILSCYSCPSYAKDLMK 1020
QY 1021 VLOGVNGEVLISQLLPMAEOLLEKIQKEPTAVLKDEAMVHLTLGKXNEFSVSLNEDPK 1080
DB 1021 VLOGVNGEVLISQLLPMAEOLLEKIQKEPTAVLKDEAMVHLTLGKXNEFSVSLNEDPK 1080
QY 1081 SLDIFIKAVHTTKELVAGMPTIQTALAKITKPPFAISDEKYOQKILRMLPDLVNCN 1140
DB 1081 SLDIFIKAVHTTKELVAGMPTIQTALAKITKPPFAISDEKYOQKILRMLPDLVNCN 1140
QY 1141 SHCAQTVSSVFKGISVNAEQVRILEPDPKAPLGTVQQRROQMOQKSGQDESVOEG 1200
DB 1141 SHCAQTVSSVFKGISVNAEQVRILEPDPKAPLGTVQQRROQMOQKSGQDESVOEG 1200
QY 1201 GSYWQVTLILLELOHKKULRSPOILLVPTLNLISRCLEPLPOEGNMETKOLILSCLL 1260
DB 1201 GSYWQVTLILLELOHKKULRSPOILLVPTLNLISRCLEPLPOEGNMETKOLILSCLL 1260
QY 1261 NIOQKLSPDGDKIPKILDEKENVELIYOCIRLSEPOTHHALLLGVAGIFPKVL 1320
DB 1261 NIOQKLSPDGDKIPKILDEKENVELIYOCIRLSEPOTHHALLLGVAGIFPKVL 1320
QY 1321 HNIMSIPTFMGAVNMRDDTYSFOVINKTVKQVPIALIQSDSGDSIEVSINVEIIVKII 1380
DB 1321 HNIMSIPTFMGAVNMRDDTYSFOVINKTVKQVPIALIQSDSGDSIEVSINVEIIVKII 1380
QY 1381 SVFVDALPHVPEHRRILPILVOLVDTGAEKFLWILILFEOYVYTKVLAAYGKDAIL 1440
DB 1381 SVFVDALPHVPEHRRILPILVOLVDTGAEKFLWILILFEOYVYTKVLAAYGKDAIL 1440
QY 1441 EADTEPFWSCEPSVOHOIOSLNILOYLLKLPKEKEETIPRAVSPKSESQEMLOVF 1500
DB 1441 EADTEPFWSCEPSVOHOIOSLNILOYLLKLPKEKEETIPRAVSPKSESQEMLOVF 1500
QY 1501 NVEHTSKQLRHFKFLSVSFMSQLSSNNFLKRVESGGEIILKGLBERLLETVLGYISA 1560
DB 1501 NVEHTSKQLRHFKFLSVSFMSQLSSNNFLKRVESGGEIILKGLBERLLETVLGYISA 1560
QY 1561 VAOSMERADKLTVKTRALLSKQYDLDBKVNALLPETETIPYIRGLVGNPLDSVRKAL 1620
DB 1561 VAOSMERADKLTVKTRALLSKQYDLDBKVNALLPETETIPYIRGLVGNPLDSVRKAL 1620
QY 1621 DLNNKLOQNIIMKKTITVTFELKVPDLAIYORKKKEGEEOAIRQTLATYLKILCKN 1680
DB 1621 DLNNKLOQNIIMKKTITVTFELKVPDLAIYORKKKEGEEOAIRQTLATYLKILCKN 1680
QY 1681 FGAENDDPFPVILXTAVKLIAPERKEKNVLSGALLCIAEVTSTLEALAIPOPLSLMPSL 1740
DB 1681 FGAENDDPFPVILXTAVKLIAPERKEKNVLSGALLCIAEVTSTLEALAIPOPLSLMPSL 1740
QY 1741 LTTMKNTSELVSSEVYLLSLAALOKVETLPHFISYLEGIIISOVTHLEKITSKMSAS 1800
DB 1741 LTTMKNTSELVSSEVYLLSLAALOKVETLPHFISYLEGIIISOVTHLEKITSKMSAS 1800
QY 1801 QANIRLSTLKKTLATTLAPRVLPAIKTYKOIEKNWKNMGPFMSILOHIGMKKKEEL 1860
DB 1801 QANIRLSTLKKTLATTLAPRVLPAIKTYKOIEKNWKNMGPFMSILOHIGMKKKEEL 1860
QY 1861 TSHOSQITAFLEALDPFRAQHSNDLEEVGKTENCIIDCLVAMVYKLSVTFPRPLPKLF 1920
DB 1861 TSHOSQITAFLEALDPFRAQHSNDLEEVGKTENCIIDCLVAMVYKLSVTFPRPLPKLF 1920


```

Db      393  OSYRRLDSVIOERKGGKLDNLINTL-----QDKKSSSTSKOREVL----- 436
Qy      433  ESSYPTLVDVLEBHKKEIADLKQKELFHQFVSLSTSGCKYQGLADSDTSLMNLPLA 492
Db      437  -----LKKISEIDQSISFEOGLAVADSA-----ADDSVVFISL----- 470
Qy      493  PVRILAMNLKIMKTISKEGVDESFIKEAVLARLDGNDIVLSAISAFEIFKHSSESSEV 552
Db      471  -----LSKRG-DKIPFLFCIA-----NSGERI 492
Qy      553  TISNLMLFORAELSKNGEWEYVULKIAADILIKELISENDQLSNOYVCLLPVY--IN 610
Db      493  IILSLIEL-----RKTIENKDDVDYOI--LPLVLSIQ 524
Qy      611  NDDTESAEKMIATYLSKSGICSLHPLRLGWEALENVTKSTKQKGLIGVANQKIEILLAD 670
Db      525  SKOTVRSRALNLTIL?-----FLEIRNE 547
Qy      671  NINIGDPSSMLKQVVDLISVGEESFNLKOKVTFHVILSVLVSCSSLKETHPEFALRVF 730
Db      548  NL-----EFSIIYGMDNDN----- 562
Qy      731  SLQKIKIKLESYITTAVEIFSEWHIELMDRGIPVEIMAHYVEINSTQKRVANEDSVELV 790
Db      563  -----KNLRMLSPVET-----KYCCSDLLDRSSEIGL-----DGYLYF 596
Qy      791  FSLKCFIYALKAPKSPFGKGDIMWNPQEKEDSRDYHLILGLFEMMLNGADAVHFRVLMK 850
Db      597  SYIPERLFTKPKKRNASK-----LAAYT 619
Qy      851  LFTKVLHEDVFOLEFKFCSVLMYTGSSLSNPLNCSVKTVLOTQALYGCAMLSQKQCKH 910
Db      620  SFLSSH-----AACSXYLSN----- 633
Qy      911  QLASISPVVTSLLINSGSVKVRARAIOCLQALSGVASEPFLIIDLHLSKKEITSDA 970
Db      634  -----VRULLLELTRVHG-----KED----- 651
Qy      971  AYYIOTLATLFEELQREKELKSHOKLSETLKNLSCVYSCPSYIANKMLKVLQGVNGEMV 1030
Db      652  -----AKQOILLRLEQSEFENS--EKFKTVSKREVEALVNCFNHTS--FTSLSLFSLSNIV 704
Qy      1031  LSQLLPMAEOLLEKIQKEPTAVLKDQEMVHLITGKTNESVSLINDDPSLDIFIRAYH 1090
Db      705  LSOAI--CRRIVE-IQS-----HL-----KDPQRLR-FVRAVI 723
Qy      1091  TTTELVAGMPTIQTALTEKITKPFPAISDEKVOQKLLRMLFDLLVNCXSHCAQTIVSSV 1150
Db      734  SODE-----QHYVVVDVDSIKIP-----DIV 755
Qy      1151  FKGISVNAEOVRLEBPBPRAKPLGTVQOKRQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1210
Db      756  FK-----KLIGSVRLVKEKPAIAKRRKIDS--HIFGDVQORLTRI 794
Qy      1211  LELLQHKKIKRSPQIIVPTLPLNLSCLEPPLPOQGNMETKQOILSCLINI-----CQK 1265
Db      795  LELLLETKQAASYPPLTASP-LFEVLINSVI-ALKEIIVSSNVLQOLLGLULLEMIGASPTTE 852
Qy      1266  LSPDGKIPKDIIDEEKFNEELIYQCIKRLSEMPOTHHALLLLLTGTVAGIPPDKVLINIMS 1335
Db      853  LSP-----SIRIDTVLVCIRSTNNPQIQKALLVSAALNANPEALHGVMP 899
Qy      1326  IFTFMGANVMELDVTYSFQVINKTVKQVIALIO--SDSGDSISVSNNVEIIVKIIISVF 1383
Db      900  IFTFMGSTVLSRDAFISIHVEQVTKVYSALIRLIGDPOSSL-----LVSCF 947
Qy      1384  VDALEPHPHERRRLIIVQVUDTGAKEKELMILLLEFGYVTKTVLAAAGKQALIEAD 1443
Db      948  VNAPPHIPQHRRLRLVRLVLTQTSNRLSVALL--QFKEKMLLA-----KSTNVVAI 998
Qy      1444  TEFVFSVCCERSVOHOIOSLNNIIOYLKLPEEK--EETIPKAVSFNKSQOEMLQOVN 1501

```

```

Db      999  HDPCLTIVQSFVADRIGSINQCSRFCLKSLEBQNSDSNGCAVLIKLDEL?-----MD 1053
Qy      1502  VETHTSKOLRHFKFLSVSFMQSLSSNNP--LKKRVESGGPEILKGEERLLETVLGYI 1558
Db      1054  VDLATLIGSLR--VKVEL--ISLVSKAKNPAPPLAMIVMSVDSFVE-IQAGLFESIKLII 1109
Qy      1559  SAVAOS--MERNADKLYVFKFRALLSKAYDLDKYNALLPTEFTIPYIRGLVGNPLREPV 1615
Db      1110  TLSQOOSNEME-----LGHVYVALRSVIVHLLPNELFCTVYLGKILHERALL 1155
Qy      1616  RRKALDLNNKLOONISMKKTIIVTRFKLVPPLALIVQKKEGGEEOAINQFALYTLK 1675
Db      1156  RRKALSTVQORVOQG-----SKVSAITALLIP--VTINISVSBEE--TQOLAMDCIA 1204
Qy      1676  LCKNFGAENPDPPFVPLTAVAKLIAPE--RKEKNVIGSALLCTAEVSTLEALAIPO 1732
Db      1205  VAKRPSA--SPELFI--SPLEAVSGPYGLKNSARDVQSAIVCTVLTNTLAARILPY 1259
Qy      1733  LPSLNPSSLITMKNTSELVSEVYLLSALALQKVEIPLPHFISPYBEGILSOVHLEKI 1792
Db      1260  LADIVYSLISLIDDAKQDEGDLLELACFSMMIDFFKVLPEFSSTIVEFTICALASDR- 1318
Qy      1793  TSEMGASQANIRLNSLAKKTATTLAPRVILPAIKKTYQIERNKNHNGPMSILQEH 1852
Db      1319  -----AFEDHAIIGELPFTIANFIPTRILMKSIFRAWPECARLSTALRLLEIBAL 1372
Qy      1853  GKMKKEELTSHOSQLTAFPLEALDPPAQSSENDLEBVGKTENCIIDCLVAMVVKISEVTF 1912
Db      1373  QNSSRGAIGTVKYSIFKFLDPSDR--RSLFPADVDAVEAOAVNVLFKFWKLSDTTF 1430
Qy      1913  RPLPFLPDMAKTEBAPXD-----RLTFYVLADCIAEKLGFLTFAGHLVKFPADTL 1966
Db      1431  RPLFHLHSMALDEDLEYTPDPSGIVSRQTFYFNLITFLDTLSYITNYAYVLD--DII 1487
Qy      1967  XOVNISKTDEAFEDSENDEKCCLLLOFLINCLYKIFLEDTQHF--ISKERAYALMPLVD 2025
Db      1488  ELLSKDT-----NSB-----VRHLVNSSLVSAFENDTEEFMVPVAPFKISPVLE 1534
Qy      2026  QLENRLGEEKQOERTKILICIAQF--SVAMADDSIMKPLVYQILKTRDSPKVRFAA 2084
Db      1535  QIQYAPLDDKY-----LVAAIVELASVASSSDN-FPSNMTOQLQYRSSVINARLIA 1586
Qy      2085  LITVLALEKLEKNYIVLLPESIPFLAELMEDECEVEHQCKITQOLETVLG--EPLOS 2142
Db      1587  IQIQTOVLGRLEGNWISTLPQSVPFIAELMEDDDQVETATLAVRIIIDRIGENSLQD 1646
Qy      2143  Y 2143
Db      1647  Y 1647

```

```

RESULT 2
553378
Probable membrane protein YJL109c - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein J0808
C/Species: Saccharomyces cerevisiae
C/Date: 05-May-1995 #sequence revision 01-Sep-1995 #text_change 09-Jul-2004
C/Accession: S53378; S56887; S57359
R/Rasmussen, S.W.
Submitted to the EMBL Data Library, February 1995
A/Description: A 37.5 kb region of yeast chromosome X includes the SWE1, MEF2, GSH1 and C
A/Reference number: S53376
A/Accession: S53378
A/Molecule type: DNA
A/Residues: 1-1769 <RAS>
A/Cross-references: UNIPROT:P42945; UNIPARC:UPI0000052F40; EMBL:X85021; NID:G728698; PID:
R/Rasmussen, S.W.
Submitted to the Protein Sequence Database, September 1995
A/Reference number: S56876
A/Accession: S56887
A/Molecule type: DNA
A/Residues: 1-1769 <RAS>
A/Cross-references: UNIPARC:UPI0000052F40; EMBL:Z49384; NID:G1008292; PID:G1008293; MIFS:

```

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:41:34 ; Search time 219.344 Seconds
(without alignments)
9041.668 Million cell updates/sec

Title: US-09-603-665-5
Perfect score: 10803
Sequence: 1 MSLAQQQLRALPQSDASL.....CCKTQQLEWVGEPIQSYF 2144

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues
Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 7.2: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10791	99.9	2144	1	HEATR_HUMAN
2	10342.5	95.7	2063	2	OST307_HUMAN
3	5807.5	53.8	2159	2	Q7S148_BRARE
4	5562	51.5	1106	2	Q6P197_HUMAN
5	4944	45.8	2288	2	Q4T723_TETNG
6	4891	45.3	1180	2	Q3UNW7_MOUSE
7	4729	43.8	958	1	HEATR_MACRA
8	4516	41.8	1090	2	Q3TNE1_MOUSE
9	4494	41.6	897	2	Q8N7L7_HUMAN
10	4072	37.7	984	2	Q3TKW0_MOUSE
11	3935.5	36.4	1336	2	Q7T152_BRARE
12	3669.5	34.0	1378	2	Q7T153_BRARE
13	3315.5	30.7	743	2	Q3V1X6_MOUSE
14	3099	28.7	733	2	Q8BLJ4_MOUSE
15	3085	28.6	733	2	Q3V1X6_MOUSE
16	2156.5	20.0	2104	2	Q3V1X6_MOUSE
17	2051.5	19.0	2096	1	HEATR_DROME
18	2051.5	19.0	2096	2	Q6AMS0_DROME
19	1926	17.8	408	2	Q8CTS5_MOUSE
20	1779	16.5	349	2	Q9E6S5_HUMAN
21	1602	15.6	349	2	Q8VCK1_MOUSE
22	1462.5	13.5	1590	2	Q8VCK1_MOUSE
23	1415	13.1	2237	2	Q54ML4_DICDI
24	1252	11.6	2237	2	Q5CAF8_ORYSA
25	1194.5	11.1	1830	1	HEATR_ARATH
26	1142	10.6	2021	2	Q5KBO2_CRYNE
27	1134	10.5	2021	2	Q5KBO2_CRYNE
28	1127.5	10.4	1801	2	Q5SNB7_CRYNE
29	1127	10.4	1857	2	Q5B1X5_EKEMI
30	1104	10.2	1802	2	Q2UIC6_DEBNA
31	1088	10.1	1798	2	Q4WLI9_ASPFU

32	1056.5	9.8	1818	2	Q592X6_CANAL	Q592X6 candida alb
33	1049.5	9.7	2251	2	Q4P937_USTWA	Q4P937 ustilago ma
34	1030.5	9.5	1788	2	Q7R2M8_NEUCR	Q7R2M8 neurospora
35	1006.5	9.3	1770	2	Q6FR93_CANCA	Q6FR93 candida gla
36	1003.5	9.3	1774	2	Q6CJ57_KLUUA	Q6CJ57 kluyveromyc
37	990.5	9.2	1649	1	YG44_SCHPO	YG44 schizosacch
38	965	8.9	1806	2	Q41696_GIBZE	Q41696 gibberella
39	956	8.8	1769	1	UTP10_YEAST	UTP10 yeast
40	950.5	8.8	1774	2	Q754J8_ASHGO	Q754J8 aschya goss
41	903	8.4	1635	2	Q6C457_YARLI	Q6C457 yarrowia li
42	870.5	8.1	1660	2	Q61Y68_CABER	Q61Y68 caenorhabdi
43	857	7.9	1650	1	HEATR_CABEI	HEATR caenorhabdi
44	852	7.9	168	2	Q2KIF9_BOVIN	Q2KIF9 bos laurus
45	773	7.2	2034	2	Q519S9_ENTHI	Q519S9 entamoeba h

ALIGNMENTS

RESULT 1
HEATR_HUMAN STANDARD; PRT; 2144 AA.
ID HEATR_HUMAN
AC Q5H583; OST308; Q9NM23;
DT 01-JUN-2001 Integrated into UniProtKB/Swiss-Prot.
DT 19-JUL-2005 sequence version 3.
DE HEAT repeat-containing protein 1 (Protein BAP28).
GN Name=HEATK1; Synonyms=BAP28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE, AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
RP GLY-2017.
RA Bouguetel et al., Chumakov I., Berry C., Cohen-Akenine A.;
RT "A novel BAP28 gene and protein."
RL Patent number WO0100669, 04-JAN-2001.
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RG Human chromosome 1 international sequencing consortium;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 1777-2144.
RX PubMed:14702039; DOI:10.1038/ng1285.
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Sekine N., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa M., Omura Y.,
RA Abe K., Kamihara K., Katsura N., Sato K., Tanigami M., Yamazaki M.,
RA Niimiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tani H., Kimura M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanohori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,
RA Togashi S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Mushino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,

RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs";
RL Nat. Genet. 36:40-45 (2004).
RN (4)
RP IDENTIFICATION BY MASS SPECTROMETRY, AND SUBCELLULAR LOCATION.
RX MEDLINE=2231727; PubMed=12429849; DOI=10.1091/mbc.E02-05-0271.
RA Scherl A., Coute Y., Deon C., Calle A., Kindbelter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RL "Functional proteomic analysis of human nucleolus";
RL Mol. Biol. Cell 13:4100-4109 (2002).
CC INTERACTION:
CC 066W87:KXN1; Inact=EBI-373098, EBI-372406;
CC -1- SUBCELLULAR LOCATION: Nuclear, nucleolar.
CC -1- SIMILARITY: Belongs to the HEATR1 family.
CC -1- SIMILARITY: Contains 1 HEAT repeat.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
CC EMBL: AX067150; CAC26776.1; -; Unassigned DNA.
CC EMBL: AL359921; CA133775.1; -; Genomic DNA.
CC EMBL: AK001221; BAA91564.1; ALT_INIT; tRNA.
CC INACC: Q9H583;
CC SWISS-2DPAGE: Q9H583; HUMAN.
CC Ensembl: ENSG00000119285; Homo sapiens.
CC HGNC: HGNC:25517; HEATR1.
CC GO: GO:0005515; F:protein binding; IPI.
CC InterPro: IPR012954; BP28 C.
CC InterPro: IPR00357; HEAT.
CC Pfam: PF08146; BP28CT; 1.
CC Pfam: PF02988; HEAT; 4.
CC PROSITE: PS50077; HEAT_REPEAT; FALSE_NEG.
CC Nuclear protein; Polymorphism.
KW CHAIN 1 2144
FT REPEAT 2106 2142 /FTid=PRO_0000186201.
FT VARIANT 1694 1694 N -> S.
FT VARIANT 1854 1854 /FTid=VAR_010939.
FT VARIANT 1967 1967 V -> A.
FT VARIANT 2017 2017 /FTid=VAR_010940.
FT VARIANT 2017 2017 N -> D.
FT VARIANT 2017 2017 /FTid=VAR_010941.
FT VARIANT 2017 2017 E -> G.
FT VARIANT 2017 2017 /FTid=VAR_010942.
FT VARIANT 2017 2017 H -> R (in Ref. 1).
FT VARIANT 2017 2017 M -> V (in Ref. 1).
SQ SEQUENCE 2144 AA; 242370 MW; 7590EBA17A5FF39D CRC64;
Query Match 99.9%; Score 10791; DB 1; Length 2144;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2138; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 MTSLAQQLORLALRQSPASLSRDEVASLSLFDPKKAATIDRDTFAIGCTGLEELGIDP 60
DB 1 MTSLAQQLORLALRQSPASLSRDEVASLSLFDPKKAATIDRDTFAIGCTGLEELGIDP 60
QY 61 SFEQFEAPLPSQALKTLERSVOTKAVNKKDENTISLFIHLSPYFLLPKPAQCLEMLIHR 120
DB 61 SFEQFEAPLPSQALKTLERSVOTKAVNKKDENTISLFIHLSPYFLLPKPAQCLEMLIHR 120
QY 121 FHIHLNODSLIAACVLPYHETRIFVRVQLKINNSKRMFWLLPVKOSGVPLAKGTLIT 180
DB 121 FHIHLNODSLIAACVLPYHETRIFVRVQLKINNSKRMFWLLPVKOSGVPLAKGTLIT 180
QY 121 FHIHLNODSLIAACVLPYHETRIFVRVQLKINNSKRMFWLLPVKOSGVPLAKGTLIT 180
DB 121 FHIHLNODSLIAACVLPYHETRIFVRVQLKINNSKRMFWLLPVKOSGVPLAKGTLIT 180
QY 181 HCYKDLGFMDPFCGLVLTYSKVFAPVPGSSAQLRVLLAFVASTIVSALVAEDVSDNIIA 240
DB 181 HCYKDLGFMDPFCGLVLTYSKVFAPVPGSSAQLRVLLAFVASTIVSALVAEDVSDNIIA 240
QY 241 KLFPIYIQGLKSLSPDTRAAATYMIICQISVAYTMENTFVNSLASQIIKTLTKISLSLKD 300
DB 241 KLFPIYIQGLKSLSPDTRAAATYMIICQISVAYTMENTFVNSLASQIIKTLTKISLSLKD 300
QY 241 KLFPIYIQGLKSLSPDTRAAATYMIICQISVAYTMENTFVNSLASQIIKTLTKISLSLKD 300
DB 241 KLFPIYIQGLKSLSPDTRAAATYMIICQISVAYTMENTFVNSLASQIIKTLTKISLSLKD 300

QY 301 LSCILVLLORQKESIGKKPFPPLCNVPPDLITLHGISSEYDVSPLLRMYLPHLVYSIH 360
DB 301 LSCILVLLORQKESIGKKPFPPLCNVPPDLITLHGISSEYDVSPLLRMYLPHLVYSIH 360
QY 361 HVTGEETEGMDGOIYRHRHEALITKISLKNNDPHLLASLLFEYISYSSQOEKSNKVS 420
DB 361 HVTGEETEGMDGOIYRHRHEALITKISLKNNDPHLLASLLFEYISYSSQOEKSNKVS 420
QY 421 LNEQPLPLRLLESKYPRTLDDVLEELHKEIADLKKQELFHOFSVLSSTSGKYQFLASD 480
DB 421 LNEQPLPLRLLESKYPRTLDDVLEELHKEIADLKKQELFHOFSVLSSTSGKYQFLASD 480
QY 481 TSLMLSLNPLAVRILANMHLKKIKMKTSGEGVDESPFKAVLARIQDNDIDVLSAISA 540
DB 481 TSLMLSLNPLAVRILANMHLKKIKMKTSGEGVDESPFKAVLARIQDNDIDVLSAISA 540
QY 541 FEIFKEHFSSEVTISMLNLFORAELSKNGEYEVKIAADILKEEILSENDQLSNOV 600
DB 541 FEIFKEHFSSEVTISMLNLFORAELSKNGEYEVKIAADILKEEILSENDQLSNOV 600
QY 601 VCLLPVAVINDDTESAEMKIAIYLSKSGICSHPLRGMEALENVIKSTKPKGLIGVA 660
DB 601 VCLLPVAVINDDTESAEMKIAIYLSKSGICSHPLRGMEALENVIKSTKPKGLIGVA 660
QY 661 NQKMIELADNINLGPSSWLKNVEDLISGEESEFNKQKVFHYILSVLSCSSLKE 720
DB 661 NQKMIELADNINLGPSSWLKNVEDLISGEESEFNKQKVFHYILSVLSCSSLKE 720
QY 721 THPPFAIRVPSLQKKIKKLESVITAVEIPSEWHIELMDRGIPVELMAHYEELNSTOR 780
DB 721 THPPFAIRVPSLQKKIKKLESVITAVEIPSEWHIELMDRGIPVELMAHYEELNSTOR 780
QY 781 VAVEDSVYFLVFSIKKTIYALKAKSPKPGDIWMNPOLKEDSDYHLILGLEFMNLGA 840
DB 781 VAVEDSVYFLVFSIKKTIYALKAKSPKPGDIWMNPOLKEDSDYHLILGLEFMNLGA 840
QY 841 DAVHPRVIMKLFKYNHEDVFOLFKFCVSVMYTGSSLSNPLNCSVKTVLOTQALVYGCAM 900
DB 841 DAVHPRVIMKLFKYNHEDVFOLFKFCVSVMYTGSSLSNPLNCSVKTVLOTQALVYGCAM 900
QY 901 LSSQKQCKQLASISSPVVTSLLNLGSPVKEVRBAIQCIALSGVASPFYLIIDHLI 960
DB 901 LSSQKQCKQLASISSPVVTSLLNLGSPVKEVRBAIQCIALSGVASPFYLIIDHLI 960
QY 961 SKABEITSDAAVYIOTLATFELQREKLLKSHQKSETLKNLLSCVYSCPSYIADLMK 1020
DB 961 SKABEITSDAAVYIOTLATFELQREKLLKSHQKSETLKNLLSCVYSCPSYIADLMK 1020
QY 1021 VLOGVNGEMLVSQLPMAEQLEKIQKEPTAVLKDEAMVHLTLGKYNBSVSLNEDPK 1080
DB 1021 VLOGVNGEMLVSQLPMAEQLEKIQKEPTAVLKDEAMVHLTLGKYNBSVSLNEDPK 1080
QY 1081 SLDIFIKAVHTTKELYAGMPTIQITALEKIKTPFPFAISDEKYQOKLRLMPLDLVNCN 1140
DB 1081 SLDIFIKAVHTTKELYAGMPTIQITALEKIKTPFPFAISDEKYQOKLRLMPLDLVNCN 1140
QY 1141 SHQAQVTSYFKGISVNAEVRLELEPPDAKPLGVYQQRQKQOKSODLESVOEVG 1200
DB 1141 SHQAQVTSYFKGISVNAEVRLELEPPDAKPLGVYQQRQKQOKSODLESVOEVG 1200
QY 1201 GSYWQVRYTLLELLOHKKKLSPOILVPTLNLJSRCLBPLPOEGNMEYTKOLISCLL 1260
DB 1201 GSYWQVRYTLLELLOHKKKLSPOILVPTLNLJSRCLBPLPOEGNMEYTKOLISCLL 1260
QY 1261 NIOQKLSPPDGKIPKQILDEKFNVELIVQCIRLSEMPQHHALLLGLVAGIFPDKVL 1320
DB 1261 NIOQKLSPPDGKIPKQILDEKFNVELIVQCIRLSEMPQHHALLLGLVAGIFPDKVL 1320
QY 1321 HNIMSIPTFMGAVNMRLDPTYSFQVINKTVKMYIPALIOSDSGDSIEVSANVEIIVKII 1380
DB 1321 HNIMSIPTFMGAVNMRLDPTYSFQVINKTVKMYIPALIOSDSGDSIEVSANVEIIVKII 1380
QY 1381 SVFVDALPHVDEHRRRLPIVLQVDTLGAKEFLMILLILFEQYVTKTVLAAVGEKDAITL 1440
DB 1381 SVFVDALPHVDEHRRRLPIVLQVDTLGAKEFLMILLILFEQYVTKTVLAAVGEKDAITL 1440

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:46:09 ; Search time 37.5044 Seconds
(without alignments)
5003.834 Million cell updates/sec

Title: US-09-603-665-5
Perfect score: 10803
Sequence: 1 MTSLAQQLQRLALPQSDASL.....CCKTIQQLTETVGEPLQSYF 2144

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /EMC_Celerra_SIDS3/prodata/2/1aa/5-COMB.pep.*
2: /EMC_Celerra_SIDS3/prodata/2/1aa/6-COMB.pep.*
3: /EMC_Celerra_SIDS3/prodata/2/1aa/7-COMB.pep.*
4: /EMC_Celerra_SIDS3/prodata/2/1aa/H-COMB.pep.*
5: /EMC_Celerra_SIDS3/prodata/2/1aa/PTUS-COMB.pep.*
6: /EMC_Celerra_SIDS3/prodata/2/1aa/RE-COMB.pep.*
7: /EMC_Celerra_SIDS3/prodata/2/1aa/backfilltest.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	427.5	4.0	721	2 US-09-248-796A-15107	Sequence 15107, A
2	310	2.9	200	2 US-09-248-796A-15108	Sequence 15108, A
3	272.5	2.5	2954	2 US-09-150-867-1	Sequence 1, Appli
4	272.5	2.5	2954	3 US-09-724-584-1	Sequence 1, Appli
5	268.5	2.5	2663	2 US-09-538-092-1352	Sequence 1252, Ap
6	266	2.5	2662	2 US-09-595-684B-311	Sequence 31, Appli
7	262	2.4	2516	2 US-09-949-016-10280	Sequence 10280, A
8	262	2.4	3056	1 US-08-508-836A-8	Sequence 8, Appli
9	262	2.4	3056	1 US-08-629-001A-3	Sequence 3, Appli
10	262	2.4	3056	1 US-08-874-266-2	Sequence 2, Appli
11	262	2.4	3056	2 US-08-642-274D-3	Sequence 3, Appli
12	262	2.4	3056	2 US-08-952-127-3	Sequence 3, Appli
13	262	2.4	3056	2 US-08-952-014C-3	Sequence 2, Appli
14	262	2.4	3056	2 US-08-984-090-2	Sequence 1, Appli
15	262	2.4	3056	2 US-10-351-733-1	Sequence 3, Appli
16	260.5	2.4	3057	2 US-09-360-416-3	Sequence 3, Appli
17	260.5	2.4	3057	2 US-10-175-225-3	Sequence 12, Appli
18	259	2.4	3066	2 US-08-952-127-12	Sequence 10, Appli
19	255.5	2.4	3433	2 US-09-538-092-1136	Sequence 1136, Ap
20	255.5	2.4	3433	2 US-09-360-416-2	Sequence 2, Appli
21	255	2.4	3056	2 US-10-175-225-2	Sequence 2, Appli
22	248	2.3	1979	2 US-09-949-016-6468	Sequence 6468, Ap
23	248	2.3	2047	2 US-09-949-016-7404	Sequence 7404, Ap
24	247.5	2.3	10182	2 US-09-134-001C-3159	Sequence 3159, Ap
25	247.5	2.3	10182	2 US-08-265-967C-2	Sequence 2, Appli
26	236.5	2.2	2470	2 US-08-265-967C-2	Sequence 2, Appli

27	236.5	2.2	2470	2 US-08-305-790B-3	Sequence 3, Appli
28	235	2.2	2482	1 US-08-328-254-6	Sequence 6, Appli
29	225.5	2.1	3259	2 US-09-949-016-6507	Sequence 6507, Ap
30	224.5	2.1	3210	2 US-09-538-092-1154	Sequence 1154, Ap
31	223	2.1	3830	2 US-09-693-205A-4	Sequence 4, Appli
32	222.5	2.1	3248	1 US-08-353-700-1	Sequence 1, Appli
33	222.5	2.1	3248	5 PCT-US95-16216-1	Sequence 16, Appli
34	220	2.0	3829	2 US-09-693-205A-16	Sequence 3, Appli
35	218.5	2.0	1972	2 US-08-875-435B-3	Sequence 2, Appli
36	218	2.0	3829	2 US-09-693-205A-2	Sequence 11433, A
37	216.5	2.0	2733	2 US-09-949-016-11433	Sequence 11, Appli
38	216	2.0	3878	2 US-09-914-258-11	Sequence 87, Appli
39	214.5	2.0	2108	2 US-09-538-092-87	Sequence 1077, Ap
40	214	2.0	1960	2 US-09-538-092-1077	Sequence 10872, A
41	213	2.0	1960	2 US-09-949-016-10872	Sequence 34, Appli
42	209	1.9	4096	2 US-09-296-662-34	Sequence 1903, Ap
43	207.5	1.9	2209	2 US-10-017-754-1903	Sequence 8, Appli
44	207	1.9	1786	2 US-08-973-462-8	Sequence 4, Appli
45	206.5	1.9	1972	2 US-08-875-435B-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-248-796A-15107
Sequence 15107, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Kelch weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 15107
LENGTH: 721
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-15107

Query Match 4.0%; Score 427.5; DB 2; Length 721;
Best Local Similarity 22.3%; Pred. No. 3.7e-26;
Matches 175; Conservative 152; Mismatches 311; Indels 145; Gaps 28;

QY 1436 KDAILEADTEPWFVSCFVSQHOIQSLNMILOYILKLPBEKKEETIPKAVSFNKSESOE 1495
DB 7 KNGDSQSVFVFGVGYMKSFSABEQLAGIVAFTKLMNDIP-----LNQLEPGE 54
QY 1496 MLQVFN-----VEHTSKQLRHFKFLSVSPMSQLSSNN-----FLKK-----VVES 1537
DB 55 EFEEVLNRPVPGTTIATLGGSGLAVALRNDLQFLDETLSSBNKHELSLTKRMALVLIDD 114
QY 1538 GGPETLKGLEERLLETIVGYISAVAQSWERNADKLT-----VKFWALLSKAYDLIDKVN 1592
DB 115 EDSEVSK-----KELVADKFRALITFPALSLDFTFNHADIK-----LCSTLYSLGNML 164
QY 1593 ALPTEFPI-PVIRGLVGNPLP-----SVRRKALDLNNKLAQO--NISKWTIVTRFKL 1644
DB 165 DLPPLNYTIDSVASLDVDTLSDLSIKVARNYVAILSRKETELNVAHCDQVIE--SV 222
QY 1645 VPDLLAVQRKKKGEEEQALNRRQALTYTLKLCNFGAENPD-----FFVPLXTAVXKLI 1700
DB 223 INNLPLIKIGIKKKNVDVL--QQAVLDFTSTIVNKKFASGREFALSDVSKLISLGIV 280
QY 1701 APER---KEKNVGSALLCIAEVTSTLEALAIPLPSPILMPSLLTTMKNTSLVSSEVYL 1757
DB 281 TTDRGLNLEQPEVITIASINAIITSYVNLGVTKLIGFPKVPFPAKIMWESTNLSIGDKESAK 340

QY 1758 L---SALAALOKVVETLPHFISPEYLEGILSOVHLEKTEISEMGA----- 1799
DB 341 LAGGSVLVLSLCVIKIPAFMSTTLEAVLTLTSSLDLDHNISSVLDLIVHMDLAQVL 400
QY 1800 -SOANIRLTSKKTLATTLAPRVLLPAIKKTYKQIEKMNKNHNGPMSILQEHIGMKKE 1858
DB 401 KSLCNWMLT-----KKFY---TNDNSGNIGLFLKTLQATINMEKK 438
QY 1859 ELTSHOSQULAFLEALDPPAASHEN-----DLEEVGKTENCIDCLVMVNVLSSEVTR 1913
DB 439 QATQATLEMRWLISAFERR-QYSENDNKKEDNNTHRLSSPHGALAFVMTKLNKSPR 497
QY 1914 PLFFKFLDMA-----KTEDAPKDRLLTFYVNLADICAEKLGFLTPAGHLVKPFAPT 1965
DB 498 PLFANLVRAVAVGEGATLKTNEV--SRLLAFRRFPKLDDELKSIITSFYSVLDDTSL 555
QY 1966 LKOVNISKTDPAFPESENDEKCCLLLOFILNCLYKIFLEPTOHFISKE-RAAXALMPLV 2024
DB 556 L-----KRFSEGSVLAVTN-----LRRILIGLTSFXYDODDYWSQOGRFSDISCP 603
QY 2025 DQLENRLGSEKFOEVRVTHLIPCIAPGVAMADSLMKPLVNYQI---LLKTRDSSPKVR 2081
DB 604 SOLSN-----IEDSIGKYLKSVSTFTVDVSSDEYNETVHELKITYSNANENSAAK 656
QY 2082 PAALITVLALAEKLEKNYIVLLPESIPFLAEMEDECEVEHOCCK-TTQOLETVLGEPL 2140
DB 657 IWSIRLKTIFQKMGQWLSYPLTVPIAELLEDDDEVEHVRGVLVIEVNLGEP 716
QY 2141 QSY 2143
DB 717 DRY 719

RESULT 2
US-09-248-796A-15108
Sequence 15108, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT FILING DATE: 1999-02-12
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/096,409
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 15108
LENGTH: 200
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-15108

Query Match 2.9%; Score 310; DB 2; Length 200;
Best Local Similarity 38.0%; Pred. No. 2,4e-17;
Matches 76; Conservative 36; Mismatches 70; Indels 18; Gaps 5;

QY 1 MTSLAQQLORL-----ALPQSDASLT-SRDEVASLTFDPKEAATIDBDTAPAGCTGL 52
DB 4 MSSLASQDSINIEKTRASVALDRKORSKLSHR---SLTFDPKQATQDVEYIETATBGL 59
QY 53 BEILGIDPSFEQFEAPLFSQAKTLERSVOTKAVNKQDENISLFLIHLSPYFLKPAQ 112
DB 60 EDLCEDDSFNKFKLTLFSETSVNLDNRNQTQKDVISQLDKNIDAFLLVGGPYGLTSLX 119
QY 113 CLEWLIHRHILHYNODSLIACVLPYHETRIFFRVYQLKINKSKRMWMLPVK-QSSV 171
DB 120 AVEWLVRRHANIHNAELMILTALPYGQHPVFEVAVLVIKQNLPOIFEMLVGYDQDLKT 179
QY 172 PLAKGTLITHCYKDLGFMDF 191

DB 180 PPASSIL-----KGMSSAMNF 194

RESULT 3
US-09-150-867-1
Sequence 1, Application US/09150867
Patent No. 6645748
GENERAL INFORMATION:
APPLICANT: Wood, Kenneth W.
APPLICANT: Sakowicz, Roman
APPLICANT: Goldstein, Lawrence S.B.
APPLICANT: Cleveland, Don W.
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Plus End-directed Microtubule Motor Required for
FILE REFERENCE: 18557C-000110US
CURRENT FILING DATE: US/09/150,867
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: US 60/058,645
EARLIER FILING DATE: 1997-09-11
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 2954
TYPE: PRT
ORGANISM: Xenopus sp.
FEATURE:
OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
OTHER INFORMATION: member of the kinesin superfamily of microtubule
OTHER INFORMATION: motor proteins
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (1)..(472)
OTHER INFORMATION: kinesin like motor domain
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (473)..(2752)
OTHER INFORMATION: rod domain
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (2753)..(2954)
OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match 2.5%; Score 272.5; DB 2; Length 2954;
Best Local Similarity 18.7%; Pred. No. 3.1e-12;
Matches 415; Conservative 365; Mismatches 809; Indels 629; Gaps 96;

QY 1 MTSLAQQLORLALPQSDASLSRDEVASLTFDPKEAATIDBDTAPAGCTGLFELLGIDP 60
DB 702 MAEKANALEELA-----LMRDNFDNIIIL---ENETLKKEIA-----DLERSLKENQ 744
QY 61 SPEQFEAPLFSQAKTLERSVOTKAVN-----KQDENISLFLIHLSPYF-----LKPRA 110
DB 745 ETNERFEI-----LEKETQKEHQAQIHIIGSLKTLVNAEMYNQULBEDLETKTLLKEQ 799
QY 111 OKCLEWL-----IHRFHILHYNODSLIACVLPYHETRIFFRVYQLKINKSKRMW 160
DB 800 EIOIAELRKRADNLOKKVRNFDLSVSMGDSSEKLC-----EIRQL----- 839
QY 161 FWLPVKGSGVPLAKGTLITH-CYKDLGFM-----DFIGSLVTK 198
DB 840 -----KOS-----LSDAEAVTRDAQKESFLSSENLKERNEDTSMNYNOKEKASLFEK 890
QY 199 SVKV-FAEYPPSSAQLRVYL--AFYASTIVSALVAEDVSNI---TAKLFPYIQKGLK 251
DB 891 QLETESKSVYKMGKEDLQKELQSAFNEINYLNGLLAGKVRDLRSVELEKKVSEPSKQLE 950
QY 252 SSLPYPRAATYMIICQISVXVTMENTFVNSLASQIILKTLIPSLIKNGLSGLIYVLQRO 311
DB 951 KALEBKALENEVTCLESEYKFLPNE--VECLKNOISKASEEIMLLKQGEHSASIT----S 1005

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:52:34 ; Search time 134.107 Seconds
(without alignments)
7405.549 Million cell updates/sec

Title: US-09-603-665-5
Perfect score: 10803
Sequence: 1 MTSLAQOLRLALPQSDASL.....CCKTIQQLTVLGEPLQSYF 2144

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /EMC_Celera_SIDS3/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /EMC_Celera_SIDS3/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
3: /EMC_Celera_SIDS3/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
4: /EMC_Celera_SIDS3/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
5: /EMC_Celera_SIDS3/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
6: /EMC_Celera_SIDS3/ptodata/2/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	10799	99.9	2144	5	US-10-745-237-230	Sequence 230, App
2	10782	99.8	2144	6	US-11-124-368A-277	Sequence 277, App
3	10722	94.6	2044	6	US-11-124-368A-278	Sequence 278, App
4	9853	91.2	2036	6	US-11-124-368A-276	Sequence 276, App
5	9853	91.2	2036	6	US-11-124-368A-280	Sequence 280, App
6	9853	91.2	2036	6	US-11-124-368A-281	Sequence 281, App
7	7756.5	71.8	1569	5	US-10-450-763-45591	Sequence 45591, A
8	5026.5	46.5	1149	5	US-10-128-558-167	Sequence 167, App
9	4494	41.6	897	4	US-10-764-425-177	Sequence 177, App
10	3694	34.2	734	4	US-10-108-260A-3300	Sequence 3300, App
11	2061	19.1	408	6	US-11-124-368A-279	Sequence 279, App
12	2051.5	19.0	2096	5	US-11-745-237-36	Sequence 36, Appl
13	2051.5	19.0	2096	5	US-11-097-143-22518	Sequence 22518, A
14	1603	14.8	325	5	US-10-128-558-358	Sequence 358, App
15	1169	10.8	2122	4	US-10-437-963-189782	Sequence 189782, A
16	1086	10.1	1798	4	US-10-128-714-8388	Sequence 8388, App
17	1085	10.0	1814	4	US-10-128-714-3388	Sequence 3388, App
18	1056.5	9.8	1818	4	US-10-032-585-7616	Sequence 7616, App
19	673	6.2	952	4	US-10-424-599-24811	Sequence 24811, A
20	663	6.1	135	4	US-10-264-049-2412	Sequence 2412, App
21	512.5	4.7	463	4	US-10-425-115-295784	Sequence 295784, A
22	434	4.0	437	4	US-10-425-114-44447	Sequence 44447, A
23	379	3.5	77	3	US-09-864-761-37036	Sequence 37036, A
24	309.5	2.9	2228	5	US-10-511-096-2	Sequence 2, Appl
25	309.5	2.9	2230	5	US-10-511-096-4	Sequence 4, Appl
26	308	2.9	2250	5	US-10-511-096-6	Sequence 6, Appl
27	308	2.9	2252	5	US-10-511-096-8	Sequence 8, Appl

28	277	2.6	2668	6	US-11-124-368A-215	Sequence 215, App
29	275.5	2.6	2633	5	US-10-450-763-36864	Sequence 36864, A
30	274	2.5	2834	4	US-10-424-599-256711	Sequence 256711, A
31	273.5	2.5	2665	6	US-11-124-368A-214	Sequence 214, App
32	272.5	2.5	2954	5	US-10-650-280-1	Sequence 1, Appl
33	268.5	2.5	2663	5	US-10-723-860-749	Sequence 749, App
34	268	2.5	2543	5	US-10-828-985A-9	Sequence 9, Appl
35	267.5	2.5	2733	6	US-11-097-143-9003	Sequence 9003, App
36	265.5	2.5	2405	4	US-10-437-963-106815	Sequence 106815, App
37	265	2.5	2503	5	US-10-828-985A-11	Sequence 11, Appl
38	264.5	2.4	2568	5	US-10-828-985A-7	Sequence 7, Appl
39	262	2.4	3056	2	US-08-984-090-2	Sequence 2, Appl
40	262	2.4	3056	3	US-09-825-476-3	Sequence 3, Appl
41	262	2.4	3056	4	US-10-060-065-20	Sequence 20, Appl
42	262	2.4	3056	4	US-10-307-077-1	Sequence 1, Appl
43	262	2.4	3056	4	US-10-059-585-41	Sequence 41, Appl
44	262	2.4	3056	4	US-10-361-562-2	Sequence 2, Appl
45	262	2.4	3056	4	US-10-116-275-338	Sequence 338, App

ALIGNMENTS

RESULT 1
US-10-745-237-230
Sequence 230, Application US/10745237
Publication No. US20050227301A1
GENERAL INFORMATION:
APPLICANT: Cyclacel Limited
APPLICANT: Glover, David
APPLICANT: Bell, Graham
APPLICANT: Freney, Lisa
APPLICANT: Midgley, Carol
TITLE OF INVENTION: Cell Cycle Progression Proteins
FILE REFERENCE: P015843MO CYK
CURRENT APPLICATION NUMBER: US/10/745,237
CURRENT FILING DATE: 2003-12-23
PRIOR APPLICATION NUMBER: US 60/439,123
PRIOR FILING DATE: 2003-01-10
PRIOR APPLICATION NUMBER: US 60/468,402
PRIOR FILING DATE: 2003-05-06
NUMBER OF SEQ ID NOS: 600
SOFTWARE: Patent version 3.1
SEQ ID NO 230
LENGTH: 2144
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Q9H563
US-10-745-237-230

Query Match 99.9%; Score 10799; DB 5; Length 2144;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2140; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	MTSLAQOLRLALPQSDASLLSRDEVASLLFDPKKAATIDRTAFAGCTGLEELLGIDP	60
DB	1	MTSLAQOLRLALPQSDASLLSRDEVASLLFDPKKAATIDRTAFAGCTGLEELLGIDP	60
QY	61	SFEQFEAPLPSQAKLTERSVQTAANKKOLDENISLPLHLSPPFLPKPAQKCLEMLIHR	120
DB	61	SFEQFEAPLPSQAKLTERSVQTAANKKOLDENISLPLHLSPPFLPKPAQKCLEMLIHR	120
QY	121	FHILVNQDLSLACVLPYHETRIFFVRIQILKINSGHGFMLLPVKOSGVPLAKGTLIT	180
DB	121	FHILVNQDLSLACVLPYHETRIFFVRIQILKINSGHGFMLLPVKOSGVPLAKGTLIT	180
QY	181	HCYKIDGFMDPISLVTKSVKFAEYFGSSAQRLVLLAFYASTIVSALVAEDVSDNIIA	240
DB	181	HCYKIDGFMDPISLVTKSVKFAEYFGSSAQRLVLLAFYASTIVSALVAEDVSDNIIA	240
QY	241	KLPFYQKGLKSLDPYRATVMIICQISVYKMENTFNVSLSQIIKTLTKPSLKDQ	300
DB	241	KLPFYQKGLKSLDPYRATVMIICQISVYKMENTFNVSLSQIIKTLTKPSLKDQ	300

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:53:49 ; Search time 11.9332 Seconds
(without alignments)
4243.670 Million cell updates/sec

Title: US-09-603-665-5

Perfect score: 10803

Sequence: 1 MTSLAQOLQRLALPQSDASL.....CCKTIOQLTVGEPLQSYF 2144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 103426 segs, 23619683 residues

Total number of hits satisfying chosen parameters: 103426

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications AA New.*
1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
7: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
8: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3694	34.2	734	7	US-11-293-697-3300 Sequence 3300, App
2	235	2.2	3113	6	US-10-505-928-325 Sequence 325, App
3	207.5	1.9	2209	7	US-11-301-554-1903 Sequence 1903, App
4	205	1.9	93	6	US-10-449-902-34905 Sequence 34905, App
5	205	1.9	93	6	US-10-449-902-35765 Sequence 35765, App
6	188.5	1.7	1328	6	US-10-504-973-332 Sequence 32, App1
7	184.5	1.7	9835	6	US-10-441-571A-4496 Sequence 4496, App
8	179	1.7	1206	6	US-10-449-902-44305 Sequence 43305, App
9	175	1.6	1205	7	US-11-293-697-2967 Sequence 2967, App
10	175	1.6	1842	6	US-10-511-937-2929 Sequence 2929, App
11	173.5	1.6	1176	6	US-10-449-902-46322 Sequence 46322, App
12	171	1.6	1325	6	US-10-449-902-51794 Sequence 51794, App
13	166	1.5	1346	6	US-10-449-902-55021 Sequence 55021, App
14	163.5	1.5	2871	6	US-10-505-928-100 Sequence 100, App
15	161	1.5	1218	6	US-10-449-902-47146 Sequence 47146, App
16	159.5	1.5	2671	6	US-10-505-928-784 Sequence 784, App
17	158.5	1.5	1116	6	US-10-449-902-47106 Sequence 47106, App
18	157.5	1.5	980	7	US-11-289-102-296 Sequence 296, App
19	155	1.4	764	6	US-10-449-902-45961 Sequence 45961, App
20	152.5	1.4	1230	7	US-11-289-102-378 Sequence 378, App
21	152	1.4	748	6	US-10-449-902-53845 Sequence 53845, App
22	152	1.4	2364	7	US-11-289-102-242 Sequence 242, App
23	151.5	1.4	1368	7	US-11-283-329-166 Sequence 166, App
24	150.5	1.4	1120	6	US-10-505-928-213 Sequence 213, App
25	150	1.4	1050	6	US-10-471-571A-4624 Sequence 4624, App

26	149	1.4	1443	6	US-10-486-020-14 Sequence 14, App1
27	148.5	1.4	868	6	US-10-953-349-674 Sequence 674, App
28	148.5	1.4	870	6	US-10-953-349-673 Sequence 673, App
29	148.5	1.4	1217	6	US-10-471-571A-4942 Sequence 4942, App
30	147.5	1.4	868	6	US-10-449-902-43523 Sequence 43523, App
31	146.5	1.4	784	6	US-10-953-349-675 Sequence 675, App
32	146.5	1.4	1146	6	US-10-471-571A-4478 Sequence 4478, App
33	145.5	1.3	1056	6	US-10-505-928-225 Sequence 225, App
34	144.5	1.3	1114	6	US-10-449-902-52920 Sequence 52920, App
35	144.5	1.3	1315	6	US-10-527-411-141 Sequence 141, App
36	143	1.3	775	7	US-11-293-697-2844 Sequence 2844, App
37	142.5	1.3	1114	6	US-10-449-902-46281 Sequence 46281, App
38	142.5	1.3	1663	6	US-10-505-928-227 Sequence 227, App
39	141.5	1.3	1056	6	US-10-449-902-45516 Sequence 45516, App
40	141	1.3	1575	6	US-10-505-928-257 Sequence 257, App
41	140.5	1.3	742	6	US-10-953-349-11696 Sequence 11696, App
42	139.5	1.3	1174	6	US-10-449-902-49996 Sequence 49996, App
43	139.5	1.3	1346	6	US-10-449-902-55428 Sequence 55428, App
44	139	1.3	990	7	US-11-293-697-3896 Sequence 3896, App
45	135.5	1.3	838	6	US-10-449-902-45252 Sequence 45252, App

ALIGNMENTS

RESULT 1
US-11-293-697-3300 Application US/11293697
; Sequence 3300, App1
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length CDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3300
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-3300

Query Match 34.2% Score 3694; DB 7; Length 734;
Best Local Similarity 100.0%; Pred. No. 7.5e-206;
Matches 734; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTSLAQOLQRLALPQSDASLSRDEVASILPDKKAAITDRRTAFIGTGHELLGIDP	60
DB	1	MTSLAQOLQRLALPQSDASLSRDEVASILPDKKAAITDRRTAFIGTGHELLGIDP	60
QY	61	SPEQEAFLPQSLATLERSVOTKAVNKQDENISLFTILHSPYFLKPKQKCLEMLIHR	120
DB	61	SPEQEAFLPQSLATLERSVOTKAVNKQDENISLFTILHSPYFLKPKQKCLEMLIHR	120
QY	121	FHILHYNDSLIACVLPYHEIRIFVRVIOQLKINNSKIRKFWMLPVKOSGVPLAKGTLIT	180
DB	121	FHILHYNDSLIACVLPYHEIRIFVRVIOQLKINNSKIRKFWMLPVKOSGVPLAKGTLIT	180
QY	181	HCKYKLGFMDFICSLVTKSVVFAEYPSASQRLVLAIFYASTISALVAADVSDNIIA	240
DB	181	HCKYKLGFMDFICSLVTKSVVFAEYPSASQRLVLAIFYASTISALVAADVSDNIIA	240
QY	241	KLFPIYQKGLKSSLPDYRAATYMIICQISVKYTMENFVNSLASQIKTLTKIPSLIKDG	300
DB	241	KLFPIYQKGLKSSLPDYRAATYMIICQISVKYTMENFVNSLASQIKTLTKIPSLIKDG	300
QY	301	LSCLIVLLQKQKPSBLGKKPPHLCNVPDLITILGISTEYVSPFLRLMHLVLSIIH	360
DB	301	LSCLIVLLQKQKPSBLGKKPPHLCNVPDLITILGISTEYVSPFLRLMHLVLSIIH	360

QY 361 HTGTEEGMDGOIYKHHLEAIIITKISLKNNDHLLASLLFEEXISYSGOEMDSNKVSL 420
DB 361 HVGTEEGMDGOIYKHHLEAIIITKISLKNNDHLLASLLFEEXISYSGOEMDSNKVSL 420
QY 421 LNEOFLPLRLLSKRYRTLDVLEELKEIADLKQELFHOFVSLSTSGKYOPLADSD 480
DB 421 LNEOFLPLRLLSKRYRTLDVLEELKEIADLKQELFHOFVSLSTSGKYOPLADSD 480
QY 481 TSLMLSNHPLAPRIANNNHKKIMKTSKEGVDESFIKEAVLARLDGDDNIDVLSAISA 540
DB 481 TSLMLSNHPLAPRIANNNHKKIMKTSKEGVDESFIKEAVLARLDGDDNIDVLSAISA 540
QY 541 FEIFKEHFSSEVTISNLLNFORAEISKNGBYEVKLIAADILIKEIISENDQLSNOVV 600
DB 541 FEIFKEHFSSEVTISNLLNFORAEISKNGBYEVKLIAADILIKEIISENDQLSNOVV 600
QY 601 VCLLPVAVINDDTESAEKIAIYLSKSGICSLHPLRGWEALENVIKSTKPGKLIIVA 660
DB 601 VCLLPVAVINDDTESAEKIAIYLSKSGICSLHPLRGWEALENVIKSTKPGKLIIVA 660
QY 661 NOKMIELADNINLGDPSMLKNVEDLISYGEESFNLKOKVTFHVILSVLVSCSSLKE 720
DB 661 NOKMIELADNINLGDPSMLKNVEDLISYGEESFNLKOKVTFHVILSVLVSCSSLKE 720
QY 721 THFPFAIRVPSLLO 734
DB 721 THFPFAIRVPSLLO 734

RESULT 2
US-10-505-928-325
Sequence 325, Application US/10505928
Publication No. US20060088532A1
GENERAL INFORMATION:
APPLICANT: Ludwig Institute for Cancer Research et al.
TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
FILE REFERENCE: 28967/39178
CURRENT APPLICATION NUMBER: US/10/505,928
CURRENT FILING DATE: 2004-08-27
PRIOR APPLICATION NUMBER: US 60/363,019
PRIOR FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 866
SOFTWARE: PatentIn 3.2
SEQ ID NO 325
LENGTH: 3113
TYPE: PRT
ORGANISM: Homo sapiens
US-10-505-928-325

Query Match 2.28; Score 235; DB 6; Length 3113;
Best Local Similarity 19.08; Pred. No: 6e-05;
Matches 427; Conservative 357; Mismatches 817; Indels 646; Gaps 101;

QY 182 CYNDLGFMDPISGLVTKSVKPAEYPCSSAQLRVILAFYASTIVSALVAEDVSDNIIAK 241
DB 751 CYDDL-----HAHY-----ESLRDL-----KSDASLVTHMEDHORSILA- 785
QY 242 LFPYIOLK-----GLKSLIPDYRAATYMIICQISVKVTWMENTFVNSLASQIIKTL 290
DB 786 ---FDQOPAHNSFANIIIGSGMPSRSE-----CLLEADQSPKNS----- 824
QY 291 TKRPSLKLDSCLIVLQOKP---ESLGKKRPFHLCNVPLDILIHGISE----- 339
DB 825 ---AIIQNVDSLEFLESQOKNSDLOKO-----CE---ELVOIKGEIEENIMKAEOMH 873
QY 340 -----TYVSPPLRYMLPRLVLSIIHHVTEETEGMDGOIYKHHLEAIIITKISLKNL 392
DB 874 QSTVAETSGKISKLQEDTSAHQNV-VAETLSALENKKELQLNDKXETQAEIQELKKS 932
QY 393 DILLASLLE-----EYISYSGOEMDS-----NKVSL-----LNEQFLP 427
DB 933 NHHLEDSLKEQLLSETLSP-EKKEMSGIISLNKREIEELTQENGTLKEINASLNOEKNN 991

QY 428 LIRLESKYPRTLDVLEELKEIADLKQELFHOFVSLSTSGKYOPLA-----DSD 480
DB 992 LIOKSES-FANYID-EREKSISELSDYOKEKLIILORCEEFGNAYEDLSQRYKAAOEKN 1049
QY 481 TSLMLSNHPLAPRIANNNHKKIMKTSKEGVDESFIKEAVLARLDGDDNIDVLSAISA 535
DB 1050 SKJECCLN-----CTSICENKNELEBOLKEAFKHOEPLTKLFAFERNNML 1100
QY 536 SAISAFELFEHFSSEVTISNLLNFORAEISKNGBYEVKLIAADILIKEIILS--EN 592
DB 1101 E-----LEIYQALRSEMTDN-----QNNKSKEAG-----LKQIMTLKEKO 1138
QY 593 DQLSNOVVCLLPVAVINDDTESAEKIAIYLSKSGICSLHPLRGWEALENVIKSTK 652
DB 1139 NKKQKEV-----NDLLQENQOLMKVTKTGECNLE-----SEPIRNSKERE 1181
QY 653 PGKLGAVANO-----KMIELADNINLGDPSMLKNVE-----D 686
DB 1182 SER-----NOCNFKQMDLEVKEISLDSYNADVLQELMLRNKELKLQESKEKXECLOHE 1236
QY 687 LISV-GEESFNLKOKVTFHVILSVLVSCSSLKEVTF--PPAI-----RVFSLOK 735
DB 1237 LQIRGDLTETSNLQWQOEI--SGLKOEIDAEKYYISGPHLSTSQDNNAHLQCSLQOT 1294
QY 736 KIKKLESVITAVEISEMHIELMDRGIPVELMAHYVEELNSTQVAVADSVFLVPSLKK 795
DB 1295 TNNKLNLEKICEIILQAEKYL-----VTELNDSREECTITATRKMAEBVGK 1340
QY 796 FIYALK-----APKSPKGDIMWNPEOLK-----EDSRDYLHLIGLFEMH 836
DB 1341 LNEVKIINDSGLHGLVEIIPGEPGEQNEQHPVSLAPLDESNSYEHITLSDKEVQ 1400
QY 837 LMGADAVHFRVIMKLFIKVHLEDF-----OLFPCSVLMTYGSLSL-----SNPLN 882
DB 1401 M-----HFAELQEFKLSLOSEHKIILHDQCMSSKMSLQTYVSLKAEMLVLSNLELN 1454
QY 883 CSVKTYLOALVYGCAMLSQKOCKHOLASIS--SPVTSLLINLNGSPVKEVRAA 938
DB 1455 PGQDLVKEMQ-LGEBGLVPSLSSGCVDPSSLSLGBSSFRALLBOGD----- 1504
QY 939 IOCLQALSGVAPFVLIIDHLISKA-----EITSDAAYIODLATLFEELQRE-KK 989
DB 1505 MSLSNLNLEGAVASANQSVDEVFCCSIOEENLTKETPSAPAKGVELESELCEVYQSLK 1564
QY 990 LKSHQKLETLKMLSCVYSCPSYIAKDLMKVLOQNGE-----VYLSQLLMAEBQLEK 1044
DB 1565 LEKMSQOIMKN-----KEIOLEQLLSERQELDCLRKQYLSNEBQWOK 1611
QY 1045 I-----QKEPTAVLKDAMVLAHLTLGKYNERSVSLNEDP-----KSLD 1083
DB 1612 LISVTLNEMSKLAERKQIOLSLLEVARLQLOGLDSSRLIGITDEDAIQGRNESC 1671
QY 1084 IFIKAVHTTKELVAGMPTIQTALKEITKPPRAAISDEKVOQKLRMLFDLLVNCNSHC 1143
DB 1672 --ISKETHSE-----TTERTPGHADVHOICDXAOODL-----NNDIEXI 1708
QY 1144 AQTSSVFPGISVANNEQVRIELEPPDKAKPLGTVQOKRQKQKQKSSODLESVOEGSGY 1203
DB 1709 TETGAKPTGECSGEOSPNTYEPGEDTQGS-----SECISELSTFS- 1751
QY 1204 WQRTVLLELLOHKKLRSPQILVPTLFN-----LH-----SRCEPLPQEOGONNE--Y 1250
DB 1752 GPNALVPMQFLQNGDHNHLOKRVKETSJENRILHVIDRBRKVESLLENKEKDESKLH 1811
QY 1251 TKQILISCLLNTIQKLSPPGKIIPKDIIDDEKFNVELIYQCIIRLSEMPOTHHALLLOT 1310
DB 1812 LOEVOQMTKEIACIELEKIVGELKKNSSDLS-----KLEVFSCDHOELLQRYVT 1861
QY 1311 VAGIPDKVLAHIMISFTPMGANVRLDPTYSFOVINKTVKAVIPALILQSDSG-----DSI 1366
DB 1862 SEGLSNDELMAHDKSREDIGNVAKANDSKERFLDYVENEL-----SRIRSEKASTIEHML 1918
QY 1367 EVSRNVEEIV-----KIISVFVALPHVPEHRLPLIVOL-----VDTLGAEK 1410